8-2012

The Identification of Quantitative Trait Loci for Sucrose and Stachyose Content in Soybean Seed

Ailan Zeng
University of Arkansas, Fayetteville

THE IDENTIFICATION OF QUANTITATIVE TRAIT LOCI FOR SUCROSE AND STACHYOSE CONTENT IN SOYBEAN SEED

THE IDENTIFICATION OF QUANTITATIVE TRAIT LOCI FOR SUCROSE AND STACHYOSE CONTENT IN SOYBEAN SEED

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Crop, Soil, and Environmental Science

By

Ailan Zeng
Nanchang University
Bachelor of Science in Biological Engineering, 2007

August 2012
University of Arkansas

ABSTRACT

Sucrose is a desirable sugar in soybean seed that affects the quality and taste of soyfoods, while stachyose is a non-digestible sugar that induces flatulence in non-ruminant animals. Therefore, soybean cultivars with high sucrose and/or low stachyose would be valuable for soyfood and meal markets. The objectives of this study were to identify quantitative trait loci (QTL) or genes associated with seed sucrose and stachyose content using simple sequence repeat (SSR) or single nucleotide polymorphism (SNP) markers. A low sucrose line, MFS-553, was crossed with a high sucrose plant introduction, PI 243545, to develop the sucrose QTL mapping population. A total of 626 random SSR plus 5361 SNP markers covering 20 soybean chromosomes were used to screen the parents MFS-553 and PI 243545, 209 SSR and 2016 SNP markers were polymorphic, of which, 97 SSR markers were used to screen the $F_{2:3}$ population and 2016 SNP markers were used to screen the $F_{3:4}$ population derived from MFS-553 x PI 243545. A normal stachyose cultivar, Osage, was crossed with a low stachyose line, V99-5089, to develop F_3 mapping population for genetic study of stachyose content. A total of 34 SSR and 56 SNP markers on chromosome 10 and 11 were used to screen the parents Osage and V99-5089, 10 SSR and 18 SNP markers were polymorphic. Of these, 5 SSR and 16 SNP markers were used to screen the $F_{3:4}$ lines derived from Osage x V99-5089. For phenotyping the sugar profile, seed of $F_{2:3,}$ $F_{3:5}$ and $F_{3:6}$ lines from MFS-553 x PI 243545 and seed of $F_{3:5}$ and $F_{3:6}$ lines from Osage x V99-5089 were analyzed for sucrose and stachyose using high performance liquid chromatography (HPLC). The sucrose content in MFS-553 x PI 243545 cross fitted to a normal distribution whereas stachyose content in Osage x V99-5089 appeared to fit to a

bimodal distribution. Composite interval mapping (CIM) and multiple interval mapping (MIM) were performed to map the QTLs in the MFS-553 x PI 243545 population. Three QTLs for seed sucrose were mapped to chromosome 5, 9, and 16, explaining 46%, 10% and 8% of phenotypic variation for sucrose content, respectively. For the stachyose mapping, data revealed a major and a minor QTL on chromosome 11 and 10 explaining 81% and 11% of phenotypic variation for stachyose content, respectively. Chi-square tests further indicated that these two QTL represent two independent genes for stachyose content. Molecular markers and QTL/genes for sucrose or stachyose content identified from this study can be used for marker assisted selection in breeding soybean lines with desired sugar profile.

This thesis is approved for recommendation
to the Graduate Council.

Thesis Director:

_____

Dr. Pengyin Chen

Thesis Committee:

_____

Dr. Kristofor Brye

_____

Dr. Edward Gbur

_____

Dr. Vibha Srivastava

THESIS DUPLICATION RELEASE

I hereby authorize the University of Arkansas Libraries to duplicate this thesis when needed for research and/or scholarship.

Agreed

_____

Ailan Zeng

Refused

_____

Ailan Zeng

ACKNOWLEGEMENTS

I would like to thank my academic advisor Dr. Pengyin Chen for his teaching, guidance and encouragement. He gave me the opportunity to study at University of Arkansas, and instructed me in the soybean breeding area with his precious experience. I would like to thank Mrs. Chen, she helped me to get adjusted to new life in USA and kindly provided me with advices in my personal life.

I would like to thank Dr. Edward Gbur, Dr. Kristofor Brye, and Dr. Vibha Srivastava for serving as my committee members and thank their guidance on my project.

I would like to thank all past and present members of soybean breeding program: Ainong Shi, Andrew Scaboo, Baishuang Yu, Bo Zhang, Bryan Stobaugh, Caroline Gray, David Moseley, Eshan Shakiba, Innan Cervantez, Jane Mokua, Jiao Wang, Jody Hedge, John Carlin, Leah Shelton, Lei Huang, Liliana Florez, Lindsay Mack, Luciano Juaregy, Moldir Orazaly, Sarah Paulson, Tetsuaki Ishibashi, Tina Hart, Xujun Fu, and Yoko Yoshikawa. They helped me with field experiments, lab experiments and data analysis.

I would like to thank my family members. My parents and sisters encouraged me to pursue the degree and supported me to achieve the goal in my career.

DEDICATION

This thesis is dedicated to my parents and my sisters.

TABLE OF CONTENTS

LIST OF TABLES

Table 1b. Seed stachyose content of Osage and V99-5089 evaluated across three locations and over two years.

Table 2. Chi-square test for goodness-of-fit to a two-gene model for 2010, 2011 and two years combination.

Table 3. Summary of SNP markers used in screening the F_3-derived population from Osage x V99-5089.

Table 4. Single marker analysis of variance for seed stachyose of 129 F_3 – derived population from the cross Osage x V99-5089 evaluated at Fayetteville (FAY), AR in 2010, Argentina (ARG) in 2011, and Costa Rica (CR) in 2011.

Table 5. Mean effect of SSR/SNP marker alleles on seed stachyose content in 129 F_3-derived population from Osagex V99-5089 in 2010 and 2011.

Table 6. Soil testing in Fayetteville, AR in 2012.

LIST OF FIGURES

Figure 1. Frequency distribution of seed stachyose content in populations derived from Osage (P3) x V99-5089 (P4), $F_{3:5}$ lines at Fayetteville, 2010 (A), $F_{3:6}$ lines in Argentina, 2011(B), and $F_{3:6}$ lines in Costa Rica, 2011(C).

Figure 2. Composite interval mapping for seed stachyose QTL on chromosome 10 in 129 F_3-derived lines from Osage x V99-5089 evaluated across three environments: (A) in Fayetteville in 2010; (B) in Argentina in 2011; (C) in Costa Rica in 2011; (D) combine data across 3 locations in 2 years.

Figure 3. Composite interval mapping for seed stachyose QTL on chromosome 11 in 129 F_3-derived lines from Osage x V99-5089 evaluated across three environments: (A) in Fayetteville, AR in 2010; (B) in Argentina in 2011; (C) in Costa Rica in 2011; (D) combine data across 3 locations in 2 years.

Chapter1 INTRODUCTION AND LITERATURE REVIEW

Soybean production and usage

Soybean (*Glycine max*) underwent its first domestication in China, and was introduced to United States in 1765. As an economical crop, soybean had the planting area fluctuating around 70 million hectares in the past 10 years in the United States (Soy Stats, 2012). The United States produced 3.0 billion bushels of soybeans, which accounted for 33% of the world soybean production in 2011(Soy Stats, 2012). Up to 43% (1.3 billion bushels) of soybean production in the United States was exported in 2011(Soy Stats, 2012). Soybean is processed mainly for protein and oil, while the defatted or toasted soybean is used as soybean meal to feed animals. Soybeans provide 68% of world protein meal and 28% of world vegetable oil meal in 2011(Soy Stats, 2012). Soybean food is popular with Asians and is processed to a variety of traditional Asian food including nonfermented products, such as Tofu, kori-tofu, yuba, nimame, soymilk, edamame, bean sprouts, and kinako, and fermented products, such as natto, soy sauce and miso. Soyfoods obtain more and more affection in the world because it is a natural source of isoflavones which reduce the cancer risk (US soyfoods directory, 2012).

Soybean seed composition

The soybean seed, on average, contains 40% protein, 20% oil, 35% carbohydrates, and 5% minerals. Carbohydrates are classified into soluble and insoluble sugar. The major components of soluble sugar are glucose, fructose, disaccharide sucrose, and raffinose family oligosaccharides (RFOs) including raffinose

and stachyose (Liu, 1997). Among the carbohydrates, the sucrose is the most abundant, followed by the raffinose family oligosaccharides (RFOs), of which stachyose is the principle constituent (Hymowitz and Collins, 1974).

Depending on the genetic make-up and growth environment, seed protein varies from 33.1% to 49.2%, seed oil varies from 14.5% to 23%, sucrose, raffinose and stachyose ranges from 2.5 to 8.2%, 0.1% to 0.9%, and 1.4 to 4.1% respectively (Liu, 1997), whereas both glucose and fructose account for small percentages (Hymowitz et al., 1972).

Sugar content also varies at different stage of seed development and fluctuates at different growth temperature. As the major energy provider during seed development, sucrose content was shown to decrease during early stage of seed development and increase towards seed maturity (Saldivar et al., 2011). Sucrose level decreased by 56% with 15% increase in temperature, while the stachyose content decreased slightly, even though the range within which temperature's influence on sugar content was determined by the soybean genetic make-up (Wolf et al., 1982). In addition, it has been found that sucrose content was significantly greater at cooler locations, whereas the stachyose content was genotype-dependent across different growing environments (Kumar et al., 2010). The variability of sugar allows breeders to modify the sugar profile in soybean.

Soybean sugar effects

Sucrose is the source of good taste and texture of soyfoods. Sucrose content determines the sweetness of raw vegetable soybean seed (Masuda, 1991) and the solids content of soymilk (Poysa and Woodrow, 2002).

The favorable effects of stachyose have been detected in several studies. It has been shown that increased desiccation tolerance accompanied increased stachyose content in developing soybean seed (Blackman et al., 1992). The soybean oligosaccharides, namely raffinose and stachyose, could alleviate oxidative stress and promote abnormal blood lipid level in rats which were treated with high-fat diets (Chen et al., 2010).

Nonetheless, the unfavorable effects of raffinose oligosaccharides (raffinose and stachyose) have gained lasting attention since 1917 (Kuriyama and Mendel, 1917). Raffinose and stachyose cause flatulence and diarrhea due to indigestion problems (Kuriyama and Mendel, 1917; Wiggins, 1984; Hata et al., 1991). Soybean meal supplied low metabolizable energy for nonruminant animals (Coon et al., 1990), which stemmed from the nondigestible carbohydrates, such as stachyose and raffinose, in soybean meal. The nondigestibility of stachyose and raffinose was due to the lack of the α-(1, 6)-galactosidase enzyme in nonruminant animals (Gitzelmann and Auricchio, 1965). The artificial removal of raffinose saccaride by ethanol extraction in soybean meal was reported to increase its digestibility from 52.1 to 63.3% (Leske et al., 1999). However, it would be a more stable and effective method to raise the metabolizable energy rate of soybean meal by elucidating the genetic mechanism for stachyose oligosaccharide metabolize pathway and genetically eliminating or minimizing the long-chain sugars from soybean seed.

Seed quality attributes for soyfoods

Soybean has the largest planted area of all crops and is the second most valuable export crop in the United States. However, only a small percent of US soybean varieties are appropriate for soyfood production. The limitation was attributed to the special requirements for soybean foods in appearance, taste, and nutrition (Carter, 1987). The characteristics of soybean seed determine the quality trait and yield of traditional soybean (Taria et al., 1990).

Large seed (>20 g/100 seed) were required for the production of tofu, edamame, miso, and soymilk, while small seed (<12 g/100 seed) were required for the production of natto, soy sauce and bean sprouts (Zhang et al., 2010). Seed size also had a positive impact on tofu yield (Bhardwaj et al., 1999).

Soybean foods processed from the soybean seed with high carbohydrates had better taste compared to those processed from the soybean seed with low carbohydrates (Taria et al., 1990). Soybean seed sucrose content was a significant determinant for tofu quality (Taria et al., 1990; Poysa and Woodrow, 2002), and it was negatively associated with yield of soymilk (Geater et al., 2000; Taria et al., 1990), glucono-delta-lactone (GDL) tofu, and calcium sulphate dehydrate (CS) tofu, and negatively associated with the hardness and firmness of tofu (Taria et al., 1990).

Seed stachyose was positively associated with yield of soymilk (Geater et al., 2000; Taria et al., 1990), GDL and CS tofu, and it positively influenced the hardness and firmness of tofu while negatively affected the solid content of tofu (Poysa and Woodrow, 2002). In addition, seed protein increased the hardness and firmness of tofu

(Poysa and Woodrow, 2002). Therefore, tofu producers preferred soybean with large seeds, high protein, and high sucrose (Poysa and Woodrow, 2002; Cicek, et al., 2006).

The fermentation of natto was affected by the composition of free sugars in soybean seed. The hardness of natto was negatively correlated with total sugar (Geater et al., 2000). High stachyose and low sucrose were preferred in natto production (Taria et al., 1990).

The production of food-grade soybean encountered a dilemma: the demand for food-grade soybeans increased by 3 to 5% annually (Griffis and Wiedermann, 1990), while the traditional breeding selection for desired food-grade soybeans usually took several years. The technology for molecular marker-assisted selection (MAS) accelerated the progress of selection for desired crop varieties.

Quantitative trait and quantitative trait loci

Most important agronomic traits are quantitative traits including seed yield, seed size, seed protein, oil, and sucrose. The quantitative traits exhibit continuous variation and are controlled by multiple quantitative trait loci (QTL), while most of the QTLs contribute small effect for phenotypic variation (Mackay et al., 2009). The marker loci are used to locate the QTLs. If a marker locus was linked with a QTL, the maker with a different genotype conferred a distinct mean value of the quantitative trait (Mackay et al., 2009).

Molecular marker-assisted selection

Molecular marker-assisted selection (MAS) is based on the associations between molecular markers and quantitative trait loci for agronomic and quality traits, MAS

combines the molecular technology (eg: molecular marker) and statistical methods (eg: multiple regression analysis) with traditional breeding. Unlike traditional breeding, which is dependent on phenotypic selection, MAS is stage independent in that the desired traits could be distinguished through detecting the linked molecular marker at a young plant growth stage. A notable example of MAS is that two simple sequence repeat (SSR) markers (Satt309 and Sat_168) were successfully used to screen the soybean varieties carrying soybean cyst nematode (SCN) resistance at the *rhg1* 1ocus (Cregan et al., 1999).

To uncover the genetic basis for these soybean quality traits at the molecular level, QTLs for several quality attributes were characterized. More than 1000 QTLs conferring over 90 agronomical important traits of soybean have been reported in Soybase (Grant et al., 2010).

The first soybean genetic linkage map (Keim et al., 1990) was constructed using 150 restriction fragment length polymorphism (RFLP) markers in a F_2 interspecific population, and this map consisted of 26 chromosomes, six more than the actual chromosome numbers of soybean. However, as the first-generation markers based on restriction fragment detection, RFLPs are not highly polymorphic in soybean, thus it's always perplexing to distinguish multiple DNA bands produced by RFLP markers. To overcome those defects, simple sequence repeats (SSRs or microsatellites) were used for construction of linkage maps (Akkaya et al., 1992). As the second-generation molecular markers, the heritable manner of SSR markers was co-dominant, which was favorable for detecting the DNA banding pattern (Jones et al., 2009). SSR markers were highly reproducible because they were based on polymerase chain reaction (PCR) and

gel electrophoresis, so SSR markers can be used to distinguish from two closely related individuals (Jones et al., 2009).

While RFLP and SSR markers which were dependent on a gel system were time-consuming and costly, single nucleotide polymorphism (SNP) markers were efficient and cost-effective and therefore SNP markers received considerable attention recently (Jones et al., 2009). Single DNA base difference plus insertions and deletions were referred to as single nucleotide polymorphisms (SNPs), which were abundant and widely distributed in the soybean genome (Zhu et al., 2003). The soybean had higher nucleotide diversity in non-coding DNA sequence than that in coding regions (Zhu et al., 2003).

To improve the resolution of the linkage map and fine map the QTL regions, all the molecular markers are integrated to the linkage map to narrow down the regions and further locate the candidate genes. For instance, two putative QTLs for soybean aphid resistance were identified on a SNP linkage map, and two new genes were fine mapped using SSR markers in these two regions (Jun et al., 2012).

QTL for soybean quality traits

A total of 90 agronomical traits for soybean have been studied for associated QTLs (Grant et al., 2010). The investigations on QTL for seed quality trait including seed protein content, seed oil content, seed size, seed calcium and hardness, seed phytic acid, have received considerable attention because of their nutritional values.

To be instructive in the breeding program, a specific QTL is required to be stable across environments. The QTL was environmentally stable if it was observed in two or

more years and for the average of the years, and the QTL was environmentally specific if it was observed either only in one year or only for the average of the years (Brummer et al., 1997). Alternatively, different QTLs for a specific trait could be detected in diverse mapping populations because of the distinct genetic backgrounds (Brummer et al., 1997), while the success of identification of common QTLs across the genetic populations would exclude the false QTLs and enhance the precision of mapping the QTLs.

QTL for soybean seed sucrose

The sucrose QTL was first studied by Maughan et al. (2000) using RFLPs, SSRs, RAPD (Random Amplified Polymorphic DNA) and morphological markers in an interspecific F_2 soybean population from the cross from V71-370 x PI407.162. A total of 17 markers distributed on seven different chromosomes (5, 7, 8,13,15,19, and 20) were significantly associated with sucrose content (Table 1), while the major markers which had phenotypic explanation of above 10% for sucrose content were all distributed on chromosome 20 (Maughan et al., 2000). The seven most significant markers together explained 53% of the total phenotypic variation for sucrose content (Maughan et al., 2000). However, the greatest level of phenotypic variation explained by these individual markers was 12.4% (Maughan et al., 2000). And the homologous lines with V71-370 had a higher average sucrose content that was greater (0.3% - 0.7%) than that of homologous lines carrying PI407.162 allele.

Kim et al. (2005) constructed a linkage map consisting of SSR markers and morphological markers using recombination inbred line (RIL) populations derived from

the cross between ‘Keunolkong’ and ‘Lksan10’ (Kim et al., 2005). In this study, four QTLs located on three chromosomes (2, 11, and 19) were identified for sucrose content and two QTLs were also found to be significantly associated with oligosaccharide content. The major sucrose QTL on chromosome 19 was flanked with the marker Satt278 and Satt523, which explained 21.4% of the total phenotypic variation for sucrose content, while the other two sucrose QTLs on chromosome 2 and 11, respectively, conferred minor effects (Table 1). However, this linkage map had low marker density, 1.0 per 15.9 CM. This could be attributed to the fact that only 199 SSR markers were chosen for screening polymorphism between parental genotypes and only 99 SSR markers showed polymorphism between parental genotypes.

Kim et al. (2006) constructed another linkage map with SSR markers and morphological markers using RIL populations derived from the cross of ‘Keunolkong’ and ‘Shinpaldalkong’ (Kim et al., 2006). Two QTLs located on two chromosomes (12 and 16) were significantly related to sucrose content, however, these QTLs explained a relatively low amount of phenotypic variation (<10%) for sucrose content. The problem of relatively low marker density also existed in this linkage map. The phenomenon that these markers explained relatively low levels of phenotypic variation for soybean sucrose content suggested that the sucrose content was a quantitative trait controlled by multiple genes (Maughan et al., 2000).

QTL for soybean seed stachyose

Stachyose content was reported to account for 1.4 to 4.1% of the seed on a dry-weight basis (Hymowitz et al., 1972). Of the raffinose saccharides, stachyose is the

most abundant and most undesirable nutritionally in soybean (Cristofaro et al., 1974). No significant difference in stachyose content was observed among environments, while a significant difference for stachyose content was detected among genotypes (Geater et al., 2000), indicating that genotype played a more important role in determining the stachyose content than the environment. In addition, it was shown that stachyose content was controlled by a single gene or a major QTL based on its phenotype segregation model in the populations (Sebastian et al., 2000; Skoneczka et al., 2009).

To reveal the inheritance mechanism of stachyose, low stachyose lines were selected from germplasm or developed by mutation method. LR28 was a soybean line selected from the natural soybean germplasm with variation of raffinose saccharide content, and LR28 had low raffinose saccharide content and low α-galactoside content (Sebastian et al., 2000). Crosses were made between four elite lines and LR28 to develop F_2 populations, the segregation ratio of a low raffinose saccharide line to a high raffinose saccharide line did not significantly deviated from the ratio of 1:3 in four elite × LR 28 F_2 populations, and the segregation ratio fit into the genetic model of a single recessive gene for low stachyose content (Sebastian et al., 2000). This recessive gene was designated as *stc* 1a, which conferred low raffinose saccharide content.

Furthermore, both LR-484 and LR33 were developed from mutation breeding and had heritable low raffinose saccharide content (Sebastian et al., 2000). An allelism test of LR484 vs LR28 showed that the single recessive gene *stc* 1b, which controlled low raffinose saccharide content in LR484, was allelic to *stc* 1a in LR28 (Sebastian et al., 2000). However, the decreased raffinose saccharide content in LR33 was controlled by a new recessive gene named as *mips* (Sebastian et al., 2000). The combination of *mips*

gene and *stc* 1x genes had additive effects, which reduced the raffinose saccharide content to zero (Sebastian et al., 2000). It was reported that the reduced raffinose saccharide content was due to the single base change in the *myo-inositol* 1-phospate synthase gene in LR33 (Hitz et al., 2002).

Skoneczka et al. (2009) identified the QTL for stachyose on chromosome 6 in the populations derived from PI 87013 x PI 200508 and PI 243545 x PI 200508, where this QTL was flanked by Sat_213 and Satt643. New markers were designed in the region between Sat_213 and Satt643 and a comparative genomic method was utilized. The result indicated that the low stachyose content in PI 200508 was due to 3bp deletion in the galatosyl transferase gene.

In addition, SSR marker Satt453 on chromosome 11 was reported to classify the low stachyose lines by effectively (Maroof and Buss, 2008), and this marker had 87% of selection efficiency in MAS (Maupin et al., 2011). Furthermore, Jaureguy (2009) identified three SSR markers on chromosome 10 significantly associated with soybean seed stachyose content (Table 1), and the stachyose QTL resided in the interval between Satt262 and Sat_282, which explained 48% of phenotypic variation.

Above all, the study for QTLs of sucrose and stachyose in soybean was not only limited, but also those putative QTLs identified were not further confirmed across diverse genetic backgrounds and environments. Thus, it's difficult to determine whether those putative QTLs were true or not, therefore, it is meaningful to identify more QTLs associated with sucrose and stachyose and test those previously identified QTL in different genetic backgrounds and environments. Furthermore, the biosynthetic pathway

for raffinose and stachyose from glucose involves several key enzymes, however, only two enzymes (galatosyltransferase and *myo-inositol* 1-phospate) were reported so far (Hitz et al., 2002; Skoneczka et al., 2009) to influence the stachyose content. Therefore, the genetic factors controlling the stachyose content deserve further study to uncover that whether any other single genes encoding key enzymes in the biosynthetic pathway determine stachyose content.

Association among agronomic traits

The most important agronomic trait is yield in any breeding program. The soybean lines with proper sugar profile are desirable without sacrificing the yield. It has been reported that the carbohydrates content had no impact on seed yield (Wilcox and Shibles, 2001). In addition, reduced raffinose saccharide content had no adverse impact on seed yield (Neus et al., 2005); sucrose content was shown to have a positive relationship with seed yield (Cicek et al., 2006); no relationship between stachyose content and seed yield was observed (Cicek et al., 2006). All of these studies indicated that seed yield was not negativelly affected by the selection for high sucrose or low stachyose lines.

However, there are complex interrelationships among seed quality traits. It has been shown that sucrose content was negatively correlated with stachyose content in soybean (Hymowitz et al., 1972 ; Neus et al., 2005), suggesting that it might be effective to identify the lines with low stachyose through selecting those lines with high sucrose, or at least this favorable correlation will not have a negative impact on selection. It has also been shown that the sucrose content increased at the expense of

the protein content in seed (Maughan,1994; Wilcox and Shibles, 2001). In addition, an increase in sucrose content accompanied an increase in oil content in seed (Wilcox and Shibles, 2001). It appears that by virtue of the association among seed quality traits, soybean lines with a desirable composition of seed sugar content could be obtained through indirect selection for protein and oil content without sugar analysis.

RESEARCH JUSTIFICATION AND OBJECTIVES

The soyfoods demand increased because of its nutritional and pharmaceutical effects. Sucrose determined the quality and taste of soyfoods, while stachyose caused flatulence for nonruminant animals. Developing soybean cultivars with a desired sugar profile is required by the increasing specialty soybean market.

Recent progress in identification of QTL and linked markers for important agronomical traits accelerated the breeding process and improved breeding efficiency. However, few QTL conferring soybean seed sucrose and stachyose have been reported.

The objectives of this study were to identify sucrose QTL, stachyose QTL, and associated molecular markers and confirm the previous reported sugar QTL in the current mapping populations and growing environments. QTL in this study would provide more information concerning the inheritance for sucrose and stachyose, and narrow down the QTL region controlling the sugar. The associated markers would help improve breeding efficiency for soybean seed sugar.

REFERENCES

Akkaya, M.S., A.A. Bhagwat, and P.B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

Bhardwaj, H.L., A.S. Bhagsari, J.M. Joshi, M. Rangappa, V.T. Sapra, and M.S.S. Rao. 1999. Yield and quality of soymilk and tofu made from soybean genotypes grown at four locations. Crop Sci. 39:401-405.

Blackman, S.A., R.L. Obendorf, and A.C. Leopold. 1992. Maturation proteins and sugars in desiccation tolerance of developing soybean seeds. Plant Physiol. 100:225-230.

Brummer, E.C., G.L. Graef, J. Orf, J.R. Wilcox, and R.C. Shoemaker. 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37:370-378.

Carter, T. 1987. Genetic alteration of seed size: Breeding strategies and market potential. P. 33-45. In: D.Wilkinson, (ed.) Proceedings of the Seventeenth Soybean Seed Research Conference. American Seed Trade Assoc., Washington, DC.

Chen, H., L.J. Liu, J.J. Zhu, B. Xu, and R. Li. 2010. Effect of soybean oligosaccharides on blood lipid, glucose levels and antioxidant enzymes activity in high fat rats. Food Chem. 119:1633-1636.

Cicek, M.S., P. Chen, M.A. Saghai Maroof, and G.R. Buss. 2006. Interrelationships among agronomic and seed quality traits in an interspecific soybean recombinant inbred population. Crop Sci. 46:1253-1259.

Coon, C.N., K.L. Leske, O. Akavanichan, and T.K. Cheng. 1990. Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. Poultry Sci. 69:787-793.

Cregan, P.B., J. Mudge, E.W. Fickus, D. Danesh, R. Denny, and N.D. Young. 1999. Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. Theor. Appl. Genet. 99:811-818.

Cristofaro, E., F. Mottu, and J.J. Wuharmann.1974. Involvement of the raffinose family of oligosaccharides in flatulence. Pp 313-335. In: H. L. Sipple and K. W. McNutt (eds.) Sugars in Nutrition. Academic Press. London, UK.

Geater, C.W., W.R. Fehr, and L.A. Wilson. 2000. Association of soybean seed traits with physical properties of natto. Crop Sci. 40:1529-1534.

Gitzelmann, R., and S. Auricchio. 1965. The handling of soy alpha-galactosidase by a normal and galactosemic child. Pediatrics 36:231-238.

Grant, D., R.T. Nelson, S.B. Cannon, and R.C. Shoemaker. 2010. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res. 38:D843-D846.

Geater, C.W., W.R. Fehr, and L. A. Wilson. 2000. Association of soybean seed traits with physical properties of natto. Crop Sci. 40:1529-1534.

Griffis, G., and L. Wiedermann. 1990. Marketing food-quality soybeans in Japan. American Soybean Association, P.O. Box 27300, St. Louis, MO 63141, USA.

Hata, Y., M.Yamamoto, and K. Nakajima. 1991. Effects of soybean oligosaccharides on human digestive organs: Estimation of fifty percent effective dose and maximum non-effective dose based on diarrhea. J. Clin. Biochem. Nutr.10:135-144.

Hitz, W.D., T.J. Carlson, P.S. Kerr, and S. A. Sebastian. 2002. Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiol. 128:650-660.

Hymowitz, T., F.I. Collins, J. Panczner, and W.M. Walker. 1972. Relationship between the content of oil, protein, and sugar in soybean seed. Agron. J. 64:613-616.

Hymowitz, T., and F.I. Collins. 1974. Variability of sugar content in seed of *Glycine max* (L.) Merrill and *G. soja* Sieb and Zucc. Agron. J. 66:239-240.

Jaureguy, L.M. 2009. Identification of molecular markers associated with seed size, protein, and sugar content in soybean. M.S.Thesis, University of Arkansas, Fayetteville, AR.

Jones, N., H. Ougham, H. Tomas, and I. Pasakinskiene. 2009. Markers and mapping revisited: finding your gene. New Phytol. 183:935-966.

Jun T.H., M.A.R. Mian, and A.P. Michel. 2012. Genetic mapping revealed two loci for soybean aphid resistance in PI567301B. Theor. Appl. Genet. 124:13-22.

Keim, P., B.W. Diers, T.C. Olson, and R.C. Shoemaker. 1990. RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics 126:735-742.

Kim, H.K., S.T. Kang, J.H. Cho, M.G. Choung, and D.Y. Suh. 2005. Quantitative trait loci associated with oligosaccharide and sucrose contents in soybean (*Glycine max L.*). J. Plant Bio. 48:106-112.

Kim, H.K., S.T. Kang, and K.W. Oh. 2006. Mapping of putative quantitative trait loci controlling the total oligosaccharide and sucrose content of *Glycine max* seeds. J. Plant Res. 119:533-538.

Kumar, V., A. Rani., L. Goyalt., A.K. Dixit., J.G. Manjaya., J. Dev., and M. Swamy. 2010. Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. J. Agric. Food Chem. 58:5081-5085.

Leske, K.L., and C.N. Coon. 1999. Nutrient content and protein and energy digestibilities of ethanol extracted, low α-galactoside soybean meal as compared to intact soybean meal. Poult. Sci. 78:1177-1183.

Kuriyama, S., and L.B. Mendel. 1917. The physiological behavior of raffinose. J. Biol. Chem. 31:125-147.

Liu, K. 1997. Soybeans: Chemistry, Technology, and Utilization. p.532. Chapman & Hall, NY.

Mackay, T.F., E.A. Stone, and J.F. Ayroles. 2009. The genetics of quantitative traits: challenges and prospects. Nature Reviews Genet. 10:565-577.

Maroof, M.A.S., and G.R. Buss. 2008. Low phytic acid, low stachyose, high sucrose soybean lines. U.S. Patent 12/033,830, field Feb.19, 2008.

Maughan, P.J. 1994. Molecular marker analysis of quantitative trait loci influencing seed quality characteristics in soybean [*Glycine max.* (L.) Merr.]. Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Maughan, P.J., M.A. Saghai Maroof, and G.R. Buss. 2000. Identification of quantitative trait loci controlling sucrose content in soybean (*Glycine max*). Mol. Breed. 6:105–111.

Maupin, L.M., M.L. Rosso, and K.M. Rainey. 2011. Environmental effects on soybean with modified phosphorus and sugar composition. Crop Sci. 51:642-650.

Neus, J.D., W.R. Fehr, and S.R. Schnebly. 2005. Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. Crop Sci. 45:589-592.

Poysa, M., and L. Woodrow. 2002. Stability of soybean seed composition and its effect on soymilk and tofu yield and quality. Food Res. Int. 35:337-345.

Saldivar, X., Y.J. Wang, P. Chen, and A. Hou. 2011. Changes in chemical composition during soybean seed development. Food Chem. 124:1369-1375.

Sebastian, S.A., P.S. Kerr, R.W. Pearisten, and W.D. Hitz. 2000. Soybean germplasm with novel genes for improved digestibility. p. 56-74. J.K. Drackely (ed.). In: Soy in animal nutrition. Federation of Animal Sci. Soc., Savoy, IL.

Skoneczka, J.A., M.A. Saghai Maroof, C.Shang, and G. R. Buss. 2009. Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. Crop Sci. 49:247-255.

Soy Stats. 2012. A reference guide to important soybean facts & figures. http://www.soystats.com/2012/default.htm (accessed 18 June. 2012).

Taria, H., H. Tanaka, and M. Saito. 1990. Effect of cultivar, seed size, and crop year on total and free sugar contents of domestic soybeans. J. Jap. Soc. Food Sci. and Tech. 37:203-213.

U.S. Soyfoods directory. 2012. http://www.soyfoods.com/soyfoodsdescriptions/descriptions.html. (accessed 18 June 2012).

Wiggins, H.S. 1984. Nutritional value of sugars and related compounds undigested in the small gut. Proc. Nutr. Soc. 43:69-75.

Wilcox, J.R., and R.M. Shibles. 2001. Interrelationships among seed quality attributes in soybean. Crop Sci. 41:11-14.

Wolf, R.B., J.F. Cavins., R. Kleiman., and L.T. Black. 1982. Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty Acids, amino acids and sugars. J. Am. Oil Chem. Soc. 59:230-232.

Zhang, B., P. Chen, S.L. Florez-Palacios, A. Shi, A. Hou, and T. Ishibashi. 2010. Seed quality attributes of food-grade soybeans form the U.S. and Asia. Euphytica 173:387-396.

Zhu, Y.L., Q.J. Song., D.L. Hyten, C.P. Van Tassell, L.K. Matukumalli, D.R. Grimm, S.M. Hyatt, E.W. Fickus, N.D. Young, and P.B. Cregan. 2003. Single-nucleotide polymorphisms in soybean. Genetics 163:1123-1134.

Table 1. Molecular markers pervious reported to be linked with soybean seed sucrose and stachyose content.

QTL name	Chr.†	Marker	cM‡	% V§	Marker type	Parents	Reference
Sucrose 1-1	5	A487_1	8.8	8.1	RFLP	V71-370 x PI407.162	Maughan et al. (2000)
Sucrose 1-2	8	A136_1	71.4	8.7	RFLP	V71-370 x PI407.162	Maughan et al. (2000)
Sucrose 1-3	8	T153_1	50.4	7.6	RFLP	V71-370 x PI407.162	Maughan et al. (2000)
Sucrose 1-4	20	A144_1	32.4	12.4	RFLP	V71-370 x PI407.162	Maughan et al. (2000)
Sucrose 1-5	13	A186_1	64.9	9.6	RFLP	V71-370 x PI407.162	Maughan et al. (2000)
Sucrose 1-6	19	A023_1	36.7	9.7	RFLP	V71-370 x PI407.162	Maughan et al. (2000)
Sucrose 1-7	19	B164_1	35.3	9.6	RFLP	V71-370 x PI407.162	Maughan et al. (2000)
Sucrose 1-8	19	B162_2	49.4	7.0	RFLP	V71-370 x PI407.162	Maughan et al. (2000)
Sucrose 1-9	7	GMSC514	4.5	7.3	SSR	V71-370 x PI407.162	Maughan et al. (2000)
Sucrose 1-10	15	A963_1	17.1	6.9	RFLP	V71-370 x PI407.162	Maughan et al. (2000)
Sucrose 2-1	11	Satt197	46.4	3.6	SSR	Keunolkong x Iksan10	Kim et al. (2005)
Sucrose 2-2	2	Satt546	87.2	6.4	SSR	Keunolkong x Iksan10	Kim et al. (2005)
Sucrose 2-3	19	Satt523	27.9	4.1	SSR	Keunolkong x Iksan10	Kim et al. (2005)
Sucrose 2-4	19	Satt278	31.2	17.3	SSR	Keunolkong x Iksan10	Kim et al. (2005)

Stachyose	10	Sat_282	59.4	48.0	SSR	MFL-552 x R95-1705	Jaureguy et al. (2009)
Stachyose	10	Satt173	53.4	37.0	SSR	MFL-552 x R95-1705	Jaureguy et al. (2009)
Stachyose	10	satt262	54.1	40.0	SSR	MFL-552 x R95-1705	Jaureguy et al. (2009)
Stachyose	11	Satt453	108.4	28.0	SSR	V99-5089 x Essex	Maupin et al. (2011)

† Chromosome.
‡ QTL position in centimorgan.
§ Percentage of phenotypic variation explained by the marker.

CHAPTER 2 IDENTIFICATION OF QUANTITATIVE TRAIT LOCI FOR SUCROSE CONTENT IN SOYBEAN SEED

ABSTRACT

Sucrose is a desirable sugar in soybean seed that affects the quality and taste of soyfoods. Measuring seed sucrose is costly and time consuming. The objective of this study was to identify quantitative trait loci (QTL) associated with seed sucrose content using simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers that can be used for indirect selection in breeding. A low sucrose line, MFS-553, was crossed with a high sucrose plant introduction, PI 243545, to develop an F_2-derived QTL mapping population. A total of 626 SSR primers covering 20 soybean chromosomes were used to screen the parents and 209 SSR markers were polymorphic. Of these, 97 were effective in differentiating the parental alleles in the initial mapping population consisting of 220 $F_{2:3}$ lines derived from MFS-553 x PI 243545. Subsequently, a total of 94 $F_{3:4}$ lines derived from the initial $F_{2:3}$ population were genotyped with 5361 SNP markers spanning 20 chromosomes, of which 2016 were polymorphic. Seed from $F_{2:3,}$ $F_{3:5,}$ and $F_{3:6}$ lines were analyzed for sucrose using high performance liquid chromatography (HPLC). Following the linkage map construction, composite interval mapping (CIM) and multiple interval mapping (MIM) were performed to locate sucrose QTL. Three QTLs for seed sucrose were mapped to chromosomes 5, 9, and 16, accounting for 46%, 10% and 8%, respectively, of the phenotypic variation for sucrose content. SSR and SNP markers associated with these QTLs can be used for marker assisted selection in breeding soybean lines with desired sugar profile.

INTRODUCTION

Soybean seed contains, on average, 40% protein, 20% oil, 35% carbohydrates, and 5% minerals on dry weight basis. Carbohydrates are classified into soluble and insoluble sugars. Soluble sugar includes glucose, fructose, sucrose, raffinose, and stachyose. Among these soluble sugars, sucrose is the most abundant, accounting for 2.5 to 8.2%, followed by stachysoe (1.4-4.1%) and raffinose (0.1-0.9%). Glucose and fructose are in trace amount (<1%) (Hymowitz and Collins, 1974; Liu, 1997).

Soybean is grown primarily for its protein and oil, while its carbohydrates determine the taste and quality of soyfoods and affect feeding efficiency of soy-meal. The sweetness and other sensory properties of tofu, soymilk, and other soyfoods are affected by the desirable sugars including glucose, fructose, and sucrose (Rackis, 1975). High sucrose is preferred for tofu production (Taria et al., 1990; Poysa and Woodrow, 2002), while low sucrose is required for natto production (Taria et al., 1990). Additionally, seed sucrose content is negatively associated with yield of soymilk, hardness and firmness of tofu (Taria et al., 1990). Therefore, production of soyfoods needs soybean varieties with specific sugar compositions.

Extracting and measuring soybean seed sucrose are costly and time-consuming, so identification of molecular markers associated with sucrose has received attention since genotypes could be classified into high sucrose or low sucrose categories through reading the associated molecular maker pattern. Maughan et al. (2000) identified 17 markers on seven chromosomes (5, 7, 8, 13, 15, 19 and 20) that were significantly associated with seed sucrose content, of which seven most significant markers together

explained 53% of the total phenotypic variation for sucrose content. In a separate study, Kim et al. (2005) reported four sucrose QTLs on three chromosomes (2, 11, and 19); the major sucrose QTL on chromosome 19 flanked with SSR markers Satt278 and Satt523 accounted for 21.4% of the total phenotypic variation in sucrose content, while the other two QTLs conferred a relatively minor effect on sucrose. Later, Kim et al. (2006) identified two additional sucrose QTLs on chromosomes 12 and 16, explaining up to 10% of phenotypic variation for sucrose content.

Comparing the results from two independent studies of sucrose QTL by Kim et al. (2005) and Maughan et al. (2000), it is evident that several QTL exist on chromosome 19, some of which are likely to be common QTL across different genetic backgrounds. Thus, chromosome 19 deserves further investigation for fine mapping of sucrose QTL using more SSR or SNP markers. In addition, other reported QTLs and markers need to be confirmed in different genetic backgrounds and across different environments before they can be effectively used in marker assisted breeding. The objectives of this research were to identify new QTL and/or confirm previously reported QTL for seed sucrose content using SSR and SNP markers that can be used to accelerate breeding soybean lines with high sucrose. In this study, new QTLs/markers are supposed to be found on different chromosomes or previously reported QTL may be confirmed in different genetic backgrounds.

MATERIALS AND METHODS

Parental materials

Two specialty soybean genotypes were used as parents in this study, MFS-553 and PI 243545. MFS-553 is a natto variety from Virginia derived from a cross between Essex (Smith and Camper, 1973) and Camp (Glenn Buss, personal communication). MFS-553 has purple flowers, narrow leaves, gray pubescence, and small seed (8-10g/100seeds) with yellow cotyledon and low sucrose content (3.6%). PI 243545 is a germplasm introduced from Japan (Germplasm Resources Information Network). PI 243545 has white flowers, broad leaves, tawny pubescence, and large brown seed (25-30g/100 seeds) with high sucrose content (7.2%).

Population development and field experiment

The cross MFS-553 × PI 243545 was made in the field in 2007 at the Agricultural Experiment Station, University of Arkansas in Fayetteville, AR. The F_1 plants were grown and confirmed as true hybrids using morphological markers in the field in Fayetteville, AR in 2008. The F_2 population, consisting of approximately 1,200 plants, was grown in Fayetteville, AR in 2009, and 250 random F_2 plants were harvested individually to form the mapping population. A sample of 60-100 seed from each $F_{2:3}$ line was sent to a winter nursery in Costa Rica for generation advancement. The remnant seed of each $F_{2:3}$ line were used to measure sugar content.

In the winter nursery, each $F_{2:3}$ line was grown in a 3-meter row with a 0.76-meter row spacing. From each $F_{2:3}$ row, a single plant (F_3) was randomly pulled and threshed to obtain recombinant inbred lines (RIL) for QTL mapping. Unfortunately, the remaining plants in each row were lost due to a late season flood in the Costa Rica winter nursery. Seed from the single F_3 plants pulled from the winter nursery were planted in single

rows ($F_{3:4}$ lines) in Fayetteville, AR in summer 2010. $F_{3:5}$ seed from each row were bulk-harvested in fall 2010. A total of 151 $F_{3:5}$ lines with adequate seed were grown in Costa Rica and Argentina. Plots were 3-meter long with 0.76 meter row spacing and were in a randomized complete block design with two replications. Seed from each plot were harvested separately for sugar analysis.

Soil analysis

Soybeans were planted in Fayetteville, AR in 2010, and were later sent to winter nurseries in Argentina and Costa Rica in 2011. The Fayetteville site is located in the Ozark Highlands of northwest Arkansas at an elevation above sea level of 426 m (Mongabay, 2012). The winter nursery in Argentina is near Pergamino, which is located in eastern part of the country at an elevation of 62 m (Mongabay, 2012). At the Argentina site, the soil surface texture is silt loam with a pH near 5.5 and SOM concentration of approximately 2.9% (INTA, 2012). The winter nursery in Costa Rica is near Upala in the northwestern part of the country at an elevation of 46 m (Mongabay, 2012). At the Coast Rica site, the soil surface pH ranges from 6.0 to 6.4, while the soil texture and SOM concentration are unknown.

At all three sites used to grow soybean in this study, the soil was cultivated before planting. All soybean plots were planted as 3-m-long rows with a row spacing of 0.76 m. In addition, soybeans were irrigated during the growing season at all three sites.

Soil samples were collected from top 10 cm at the four corners (A, B, C, and D) of the study area in Fayetteville, oven-dried at 70^{o}C for 48 hr, and crushed and sieved to pass a 2-mm mesh screen. Sieved soil was used for soil particle-size analysis based

on the 2-hr hydrometer method described by Arshad et al (1996). Sieved soil was also used to determine soil pH and electrical conductivity potentiometrically on a 1:2 (m/v) soil-to-water paste and Mehlich-3 extractable soil nutrients including P, K, Ca, Mg, S, Na, Fe, Mn, Zn, Cu and B (Watson et al., 1998; Donahue et al., 1983; Tucker, 1992). Soil organic matter was determined by loss-on-ignition (Combs and Nathan,1998; Wolf and Miller, 1998).

Sugar analysis

Sugar extraction followed the protocol previously described by Hou et al. (2009). Briefly, seed were ground to a fine powder using a coffee bean grinder (Krups®, Shelton, CT), then the powder was passed through a 100-µm standard testing sieve (VWR International, West Chester, PA) to obtain a homogeneous sample. Then, a 0.15-g sample was weighed and dissolved in 1.5 mL of distilled water, vortexed and then shaken horizontally at 200 rpm for 20 min, followed by centrifugation at 10000 rpm for 10 min. Then, 500 µL of the supernatant were transferred to a fresh tube. The sample was then dissolved with 700 µL of acetonitrile for 30 min. followed by centrifugation at 10000 rpm for 10 min. Then, the supernatant was filtered through a 0.2-µm filter, and finally 24 µL of sugar extract were re-dissolved with 576 µL of ddH_2O for HPLC analysis.

The HPLC system (Dionex DX500) consisted of a GS50 gradient pump, AS40 automated sampler, ED40 electrochemical detector, a Chromeleon Chromatography Management System, and a CarboPac PA 10 pellicular anion-exchange resin column. A series of sugar concentrations of 10, 20, 40, 60, and 80µg/µL were used to construct

the standard curve for glucose, fructose, sucrose, raffinose and stachyose. The sugar concentration was converted into mg/g based on dry seed matter.

DNA extraction

Tissue samples were collected from the top trifoliolate leaves of 250 randomly chosen F_2 plants at Fayetteville in summer 2009. Tissue samples were also bulk collected from each of the 151 $F_{3:4}$ lines in summer 2010. Leaf samples were stored at -80 °C until DNA extraction. Total genomic DNA was extracted using the CTAB (hexadecyltrimethyl ammonium bromide) method (Kisha et al., 1997). Briefly, tissue samples were ground in liquid nitrogen with a mortar and pestle and the extraction buffer was added to the ground sample followed by chloroform:isoamyl alcohol (24:1) to remove the protein. The DNA was dissolved by 0.1 × TE buffer and the concentration was measured using Bio-Tek PowerWave XS Microplate Spectrophotometer (BioTek Instruments, Winooski, VT). The DNA solution was stored at -80°C.

Polymerase chain reaction and SSR marker screening

Polymerase chain reaction (PCR) was performed in an iCycler ThermalCycler (Bio-Rad, Hercules, CA). The total reaction volume of PCR was 15.2 μL, consisting of 4 μL of template DNA (20ng/μL), 1μL primers (0.5 μM), 1 unit of Taq polymerase, 1.8 μL $MgCl_2$ (2.5 mM), 0.9 μL dNTP (2.5 mM), and 4.3 μL ddH_2O. The PCR program was set as 94°C for 4 min. (predenaturation) followed by 32 cycles of reaction which consisted of 94°C for 25 s (denaturation), 47°C for 25 s (annealing), 68°C for 25 s (extension), and another cycle for final extension at 72°C for 7 min.

The PCR products were loaded on the 6% non-denaturing polyacrylamide gel electrophoresis (PAGE) (C.B.S. Scientific, San Diego, CA) and the gel was run at a constant 350V for 2 hours. The PCR products were stained with ethidium bromide and were visualized under the ultraviolet light (Wang et al., 2003).

Parents were screened using 626 SSR markers randomly chosen from 20 chromosomes (Cregan et al., 1999; Song et al., 2004) (Table 3). Only polymorphic markers between the two parental genotypes were used to screen the $F_{2:3}$ lines. The F_2 individuals carrying the same allele from MFS-553 were scored as "A"; individuals carrying the same allele from PI 243545 were scored as "B", and heterozygous individuals carrying one allele from each parent were scored as "AB". The Chi-Square test was used to test for goodness-of-fit of each polymorphic marker to the expected 1A:2AB:1B ratio.

SNP marker screening

For genetic map construction, two parental and 94 $F_{3:4}$ DNA samples were genotyped with 5361 SNP markers (dbSNP-NCBI, 2012) using the Illumina Infinium® Genotyping HD BeadChip (652k SNPs) on Illumina iScan (Illumina, San Diego, CA) at the genotyping core facility of Michigan State University, East Lansing, MI. Each 4 uL sample with >200 ng/uL genomic DNA was used for SNP analysis. Intensities of the beads fluorescence were detected using the Illumina iScanTM Reader and the allele calling for each SNP locus were performed using Ilumina's BeadStudioTM software (Illumina, San Diego, CA, v3.2.23). Three genotypes were represented by the counted alleles: AA, BB and AB. (Wang D.C., 2012, unpublished).

Data analysis

The Shapiro-Wilk (*w*) statistic from JMP 9.0 (SAS Institute, Cary, NC) was used to test the normality of the sucrose content distribution for the $F_{2:3}$ lines and F_3-derived lines. Broad-sense heritability (H^2) of seed sucrose content was calculated using the following equation (1).

$$H^2=\sigma^2_g/[\sigma^2_g+ (\sigma^2_{gxe}/e) + (\sigma^2 /re)] \quad (1)$$

σ^2_g is the genetic variance, σ^2_{gxe} is the genotype by environment interaction, σ^2 is the error variance, r is the number of replicates, and e is the number of environments (location or year).

The associations between seed sucrose content and molecular markers were tested by single factor analysis of variance (ANOVA) at the 0.05 significance level using the PROC GLM procedure of SAS 9.0 (SAS Institute, Cary, NC). Linkage maps were constructed using JoinMap 4.0 (Van Ooijen, 2006) and the threshold for logarithm of odds (LOD) for linkage group construction was set as 3.0. Regression mapping of each chromosome/linkage group (LG) was performed with a Haldane mapping function (Haldane, 1919).The composite interval mapping (CIM) was performed using QTL Cartographer 2.5 (Basten et al., 1999) and Qgene (Joehanes et al., 2008). One thousand permutation with a walk speed of 1cM and experiment-wise α=0.05 was adopted to establish the empirical significance threshold (Churchill and Doerge, 1994). Multiple-interval mapping (MIM) analysis was performed to determine the optimum position and interaction of QTL. MIM was performed based on the model $c(n) = \ln(n)$

with a walk speed of 1 cM. MapChart (Voorrips, 2002) was used to create the LOD plots based on JoinMap 4.0 and QTL Cartographer 2.5 data.

RESULTS

Soil testing results

Soil testing results showed that the soil at the Fayetteville site is a Captina silt loam (fine-silty, siliceous, active, mesic Typic Fragiudult; SSS-NRCS-USDA, 2012) with 33.2 % sand, 50.3% silt, and 16.5% clay in the top 10 cm and other soil chemical properties including soil pH ranging from 6.1 to 6.3 with an average of 6.2 and soil organic matter (SOM) concentration ranging from 1.0 to 1.4% with an average of 1.2% (Table 10).

Phenotypic data

PI 243545 exhibited consistently higher sucrose content than MFS-553 by 1.7-3.8% in all environments used in this study. The sucrose differential between the high and low parents was consistent in all environments except for in Argentina in 2011. The range of sucrose contents in the mapping population exceeded the low and high parents, indicating the presence of transgressive segregation, while the mean of the population was close to the median of the two parents. Both of these parameters were typical features of a quantitative trait. The normality tests using the Sharpiro-Wilk statistic showed that the seed sucrose contents were normally distributed across locations and years (Fig. 1 and Table 1), indicating that seed sucrose content is a quantitative trait controlled by multiple genes/QTL with small effects and influenced by environmental conditions.

The ANOVA model for seed sucrose was significant with R^2 value of 0.95 indicating that a correct model was adopted in this study (Table 2). The genotype, location, and genotype x location variance were all significant. Location accounted for the largest percentage of variance, followed by the genotype (Table 2). However, data from the two locations were highly correlated and most of the genotypes in the population ranked similarly between the two locations, as also reflected in the relatively low percent of genotype x location variance component. As a result, a relatively high broad-sense heritability (H^2) estimate (0.74) was obtained for seed sucrose content based on variance components.

QTL mapping in $F_{2:3}$ population by SSR markers

Among 626 SSR markers screened, 209 (33%) were polymorphic between the parental genotypes (Table 3). These polymorphic makers covered all 20 chromosomes with 10.45 SSRs per chromosome and 11.6 cM per polymorphic marker on average. Of the 209 polymorphic markers, 97 were effective in differentiating the parental alleles in the mapping population and were used to construct a linkage map. The grouping and relative position of the markers were consistent with the public soybean linkage map, as partially shown in Figure 2.

Six SSR markers on four chromosomes (9, 17, 18, and 19) were significantly associated with sucrose content in the single marker analysis (Table 4). All six markers accounted for 3-4% sucrose variation with favorable alleles for high sucrose coming from both parents. The marker alleles on chromosomes 17 and 18 from the high sucrose parent (PI 243545) contributed positively to the sucrose content in the

population, while other sucrose-contributing alleles were from the low sucrose parent (MFS-553) and were located on chromosomes 9 and 19 (Table 4).

In the CIM analysis, two sucrose QTLs were found on chromosome 5 and 16, respectively. The sucrose QTL on chromosome 5 was flanked with Sat_344 and Sat_407, which explained 2% of the phenotypic variation for seed sucrose content (Fig. 2A), while the QTL on chromosome 16 was located between Sat_366 and Satt431, which explained 3% of the phenotypic variation for seed sucrose content (Fig. 2B).

QTL mapping in F_3-derived populations by SNP markers

A total of 5361 SNP markers randomly distributed on 20 chromosomes were used to genotype $F_{3:4}$ lines and 2016 SNP loci (38%) were polymorphic. A total of 1897 SNP markers were mapped on 20 chromosomes, representing 1283 unique SNP loci (Table 5). This linkage map covered 6602 cM with an average coverage of 5.1 cM per marker, generating a high density soybean linkage map (Table 5 and Fig. 3A-E).

In the single marker analysis, eight SNP markers on five chromosomes (5, 9, 13, 16, 17) were significantly associated with seed sucrose content in at least one of the three environments as well as in all environments combined, while two SNP markers on chromosome 18 and 19, respectively, were marginally associated with sucrose content (Table 6). The homozygous lines with the PI 243545 allele had higher sucrose content than those with MFS-553 allele on chromosome 5, 13, 16, 17, and 18 in the two years, whereas the allelic effect was opposite at the marker allele on chromosome 9 and 19 with the MFS-553 allele being positive in contributing to sucrose in both years (Table 7). On average, the increase of sucrose content contributed by the PI 243545 allele ranged

from 0.28 to 1.09%, whereas the MFS-553-marker allele on chromosome 9 and 19 increased the sucrose content by up to 0.53% (Table 7). It appears that the marker alleles on chromosomes 16 are important for sucrose content.

In the CIM analysis, the empirical significance threshold was computed as a LOD value of 2.5 in the $F_{3:4}$ mapping population in both 2010 and 2011. Based on the average seed sucrose content across three locations over two years, a major sucrose QTL was mapped on chromosome 5, flanked by ss245634120 and ss245672731, tightly linked with ss245668753 (Fig. 4A-D). This QTL was mapped at 125.0 cM on chromosome 5 and tentatively named as *Suc1*. *Suc1* had a LOD value of 5.7, significantly higher than the threshold, and explained 46% of phenotypic variation for sucrose content for the combine data. Based on the average seed sucrose content across two locations in one year, a minor sucrose QTL was mapped on chromosome 9, located on the interval between ss246785428 and ss246817520, and was tightly linked with ss246796276 (Fig. 5). The sucrose QTL on chromosome 9 accounted for 10% variation in seed sucrose content for the combine data and was tentatively designated as *Suc2*. Another QTL was identified on chromosome 16 and tentatively named as *Suc3.* This QTL was located on the interval between ss249194115 and ss248932393 and tightly linked with ss249186914. This QTL was identified in Fayetteville in 2010 and Argentina in 2011 and explained 8% of variation in seed sucrose content for the combine data.

In the MIM analysis, the optimum positions were determined to be at 125.3 cM on chromosome 5 for *Suc1* and 76.2 cM on chromosome 9 for *Suc2*, respectively. No significant epistasis interaction was detected among the sucrose QTL in this study.

DISCUSSION

The average seed sucrose contents for both parents and F_3-derived lines were lower in Costa Rica than other locations (Table 1), likely due to the higher temperature (Table 9) in Costa Rica during the growing season which resulted in the degradation of the sucrose into the glucose and fructose. This observation was in agreement with results from other studies where increased temperature reduced the sucrose level (Wolf et al., 1982; Hou et al., 2009). Evidently, the sugar profile is influenced by environmental conditions, as expected for a quantitative trait, justifying the need of multiple locations and years for sugar phenotyping in QTL mapping.

Variation among genotypes, locations, and genotype x location interaction were all significant for sucrose content in $F_{3:6}$ lines in 2011, as expected for any quantitative trait evaluations (Table 2). However, the genotype and location effects were greater than other sources of variation when comparing variance components. It is important to note that although the genotype by location interaction was statistically significant, it was almost negligent in relevance to the error term. This was also supported by the fact that lines in the mapping population ranked similarly across locations. In addition, the variation between replication within a location was also negligible relative to the error term. These results indicate that genotypic variation was most important and genotypic ranking was stable across replications and locations (Table 2). The results also support the observations in other studies where genotypic and environmental effects were significant for soybean seed sucrose (Cicek et al., 2006; Maupin et al., 2011). The heritability of sucrose in this study (0.74) was similar to that reported by Maughan et al. (2000) and Kim et al. (2005).

The genetic map in the study was constructed using $F_{3:4}$ lines and SNP markers selected from national center for biotechnology information (NCBI). This map is unique because a specific set of SNP markers were used for the DNA chip in genotyping the $F_{3:4}$ lines which have not been reported in public soybean linkage map. Therefore, it is not possible to compare the genetic positions of these SNP markers (Fig. 3) relative to the public map at this time. However, the alignment of our SNP marker/ QLT positions with the genetic positions in the public map will be realized when the information on this specific set of SNP markers becomes published.

In this study, QTL mapping was done using SSR markers in $F_{2:3}$ lines and SNP markers in the F_3–derived lines derived from the same mapping population. SSR analysis revealed that the sucrose QTL on chromosome 5 identified in $F_{2:3}$ lines resides between the interval Sat_344 and Sat_407. The physical position of this interval was between 3691696 bp and 33718372 bp in the 'Williams 82' physical map (Table 8). SNP analysis clearly showed that SNP marker ss245668753 linked to the sucrose QTL on chromosome 5 was located at 27447797 bp in the 'Williams 82' physical map (Table 8). Thus, the QTL identified by both marker systems probably are the same QTL for sucrose. These results also showed that the QTL region identified from $F_{2:3}$ lines was confirmed and narrowed down by the F_3–derived population. The confirmation of putative QTL for a trait of interest by two different marker systems and two generations is somewhat unique in this study as compared to other QTL mapping studies.

The sucrose QTL on chromosome 16 identified by SSR markers in $F_{2:3}$ lines resides between the interval Sat_366 and Satt431, corresponding to the region between 30036653 bp and 35718642 bp in the 'Williams 82' physical map. SNP analysis for the

F_3-derived lines showed that SNP marker ss249186914 linked to the sucrose QTL on chromosome 16 was located at 32854900 bp. In comparison with the reference 'Williams 82' physical map, these two QTL regions were in the similar proximity, with the sucrose QTL in F_3-derived population being narrowed down to a smaller region (Table 8). Based on the physical position of the linked SNP marker, this QTL region was closer to Satt431, making this SSR marker useful in marker-assisted selection for sucrose. In a previous study, a sucrose QTL on chromosome 16 was identified by Kim et al. (2006). This QTL was flanked by satt287 and sct_065, which was distantly located upstream form the QTL identified in the study.

Although many significant markers were identified by Maughan et al. (2000) for seed sucrose content using single marker analysis (SMA), no QTL was claimed in their population. SMA was based on the simple linear regression that only suggests trait and marker association and does not taking into consideration of the interaction effect of the marker alleles and therefore is not adequate for QTL discovery. This would make it difficult to use their reported markers in molecular breeding for sucrose. In other studies, sucrose QTL were mapped by Kim et al. (2005, 2006) using data from only one location. Obviously, these QTLs were tentative and need to be confirmed in multiple environments.

In conclusion, new sucrose QTLs were identified in this study using two sets of markers, multiple generations, and multiple environments. Identified QTL regions were confirmed by SSR and SNP markers and different generations of materials from the mapping population. These QTL and associated SNP and SSR markers could be potentially used for MAS.

REFERENCES

Arshad, M.A., B. Lowery, and B. Grossman. 1996. Physical tests for monitoring soil quality. p. 123-141. *In* J.W. Doran and A.J. Jones (ed.) Methods for assessing soil quality. SSSA Spec. Publ. 49. SSSA, Madison, WI.

Basten, C.J., B.S. Weir, and Z.B. Zeng. 1999. QTL Cartographer, version 1.13. Department of Statistics, North Carolina State University, Raleigh, NC.

Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963-971.

Cicek, M.S., P. Chen, M.A. Saghai Maroof, and G.R. Buss. 2006. Interrelationships among agronomic and seed quality traits in an interspecific soybean recombinant inbred population. Crop Sci. 46:1253-1259.

Combs, S.M. and M.V. Nathan. 1998.Recommended chemical soil test procedures for the N central region. Soil organic matter. p. 53-59. MO Agricultural Experiment Station SB 1001. Columbia, MO.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N.Kaya, T.T. Van Toai, D.G. Lohnes, J.Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464-1490.

Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. Available from: http://www.ncbi.nlm.nih.gov/SNP/. (accessed 18 June 2012).

Donahue, S.J. 1983. Determination of soil water pH. Reference soil test methods for the S region of the US. p. 1-4. S Coop. SB. 289. U of G College of Agriculture Experiment Station. Athens, GA.

Donahue, S.J. 1983. Determination of specific conductance in supernatant 1:2 soil:water solution. Reference soil test methods for the S region of the US. p. 8-14. S Coop. SB. 289. U of G College of Agriculture Experiment Station. Athens, GA.

Haldane, J.B.S. 1919. The combination of linkage values, and the calculation of distance between the loci of linked factors. J. Genet. 8:299-309.

Hou, A., P. Chen, J. Alloatti, D. Li, L. Mozzoni, B. Zhang, and A. Shi. 2009. Genetic variability of seed sugar content in worldwide soybean germplasm collections. Crop Sci. 49:903-912.

Hymowitz, T., and F.I. Collins. 1974. Variability of sugar content in seed of *Glycine max* (L.) Merrill and *G. soja* Sieb and Zucc. Agron. J. 66:239-240.

JMP, Version 9. 2010. SAS Institute Inc.,Cary, NC.

Joehanes, R., and J.C. Nelson. 2008. QGene 4.0, an extensible Java QTL-analysis platform. Bioinformatics 24: 2788-2789

Kim, H.K., S.T. Kang, J.H. Cho, M.G. Choung, and D.Y. Suh. 2005. Quantitative trait loci associated with oligosaccharide and sucrose contents in soybean (*Glycine max L*.). J. Plant Biology 48:106-112.

Kim, H.K., S.T. Kang, and K.W. Oh. 2006. Mapping of putative quantitative trait loci controlling the total oligosaccharide and sucrose content of *Glycine max* seeds. J. Plant Res. 119:533-538.

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317-1325.

Lander, E.S., and D. Bostein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Liu, K. 1997. Soybeans: Chemistry, Technology, and Utilization. p.532. Chapman & Hall, NY.

Maughan, P.J., M.A. Saghai Maroof, and G.R. Buss. 2000. Identification of quantitative trait loci controlling sucrose content in soybean (*Glycine max*). Mol. Breed. 6:105-111.

Maupin, L.M., M.L. Rosso, and K.M. Rainey. 2011. Environmental effects on soybean with modified phosphorus and sugar composition. Crop Sci. 51:642-650.

Mongabay. 2012. http://population.mongabay.com/ (accessed 17July 2012).

National Institute of Agricultural Technology (INTA), Argentina. 2012. http://anterior.inta.gov.ar/suelos/cartas/series/Pergamino.htm (accessed 17July 2012).

Rackis, J.J. 1975. Oligosaccharides of food legumes: alpha-galactosidase activity and flatus problem, in physiological effects of food carbohydrates, Jeanes, A. and J. Hodge, Ed., pp. 207-222, American chemical society, Washington, DC, USA.

Poysa, M. and L.Woodrow. 2002. Stability of soybean seed composition and its effect on soymilk and tofu yield and quality. Food Res. Int. 35:337-345.

SAS Institute. 2010. SAS/STAT Software Version 9.2. SAS Institute Inc, Cary, NC.

Smith,T.J., and H.M. Camper. 1973. Registration of ‘Essex’ soybean. Crop Sci.13:495.

Soil Survey Staff (SSS), Natural Resources Conservation Service (NRCS), United States Department of Agriculture (USDA). 2012. Official Soil Series Descriptions.

Available online at http://soils.usda.gov/technical/classification/osd/index.html. (accessed18 July, 2012).

Song, Q.J., L.F. Marek, R.C. Shoemaker, K.G. Lark, V.C. Concibido, X. Delannay, J.E. Specht, and P.B. Cregan. 2004. A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 109:122-128.

Taria, H., H. Tanaka, and M. Saito. 1990. Effect of cultivar, seed size, and crop year on total and free sugar contents of domestic soybeans. J. Jap. Soc. Food Sci. & Tech. 37:203-213.

Tucker, M.R. 1992. Determination of phosphorous by Mehlich 3 extraction. p. 6-8. In: S.J. Donohue (ed.), Soil and media diagnostic procedures for the Southern Region of the United States. Vir. Agric. Exp. Stn. Ser. Bull. 374. Blacksburg, VA.

Van Ooijen, J.W. 2006. JoinMap 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, the Netherlands.

Voorrips, R.E. 2002. MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. J. Hered. 93:77-78.

Wang, D., J. Shi, S.R. Carlson, P.B. Cregan, R.W. Ward, and B.W. Diers. 2003. A low-cost, high throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA marker. Crop Sci. 43:1828-1832.

Watson, M.E. and J.R. Brown. 1998. pH and lime requirements. N Central Regional Research Pub. 221:13-16.

Wolf, R.B., J.F. Cavins., R. Kleiman, and L.T. Black. 1982. Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty Acids, amino acids and sugars. J. Am. Oil Chem. Soc. 59:230-232.

Wolf, A.M. and R.O. Miller. 1998. Development of a NAPT program for soil and plant analysis. Commun. Soil Sci. Plant Anal. 29:1685-1690.

Zhang, B., P. Chen, C.Y. Chen, D. Wang, A. Shi, A. Hou, and T. Ishibashi. 2008. Quantitative trait loci mapping of seed hardness in soybean. Crop Sci. 48:1341-1349.

Table 1a. Seed sucrose content of parents and mapping populations (220 $F_{2:3}$ lines, 94 F_3-derived lines) from MFS-553 x PI 243545 evaluated across three locations and over three years.

Year-Loc-Population	No. lines	Mean	SD†	Range	Prob < W‡
2009 Fayetteville $F_{2:3}$	220	6.47	0.67	4.20 - 8.70	0.1964
2010 Fayetteville $F_{3:5}$	94	7.33	1.03	4.69 - 9.83	0.6220
2011 Argentina $F_{3:6}$	94	7.28	0.64	5.61 - 9.03	0.8404
2011 Costa Rica $F_{3:6}$	94	5.13	0.71	3.42 - 7.47	0.5131

† Standard deviation.
‡ Sharpiro-Wilk value for the normality test.

Table 1b. Seed sucrose content (%) of MFS-553 and PI 243545 evaluated across three locations and over two years.

	MFS-553		PI 243545	
	Mean	**SD†**	**Mean**	**SD†**
2009 Fayetteville	4.12	0.30	7.37	0.44
2010 Fayetteville	4.07	0.51	7.17	0.98
2011 Argentina	5.9	0.09	7.32	0.32
2011 Costa Rica	3.63	0.42	6.39	0.66

† Standard deviation.

Table 2. Analysis of variance for seed sucrose of 94 $F_{3:6}$ population from the cross MFS-553 x PI 243545 grown in Argentina and Costa Rica in 2011.

Source	Degree of freedom	Mean square	Variance components	P-value	R^2
Model	188	3.16	-	< 0.0001	0.95
Location	1	414.79	2.31	< 0.0001	
Replication(Location)	2	4.96	0.06	< 0.0001	
Genotype	93	1.33	0.27	< 0.0001	
Location x Genotype	92	0.31	0.07	0.0005	
Error	176	0.17	0.17		

Table 3. Summary of SSR markers used in the initial screen of the parental genotypes and $F_{2:3}$ population derived from MFS-553 x PI 243545.

Chr.†	LG‡	Length (cM)§	No. of Markers¶	Coverage#	Polymorphic markers††	Marker coverage‡‡
1	D1a	109.7	28	3.9	12	9.1
2	D1b	138.0	38	3.6	13	10.6
3	N	103.3	27	3.8	9	11.5
4	C1	132.5	27	4.9	8	16.6
5	A1	101.6	27	3.8	11	9.2
6	C2	151.9	36	4.2	10	15.2
7	M	140.7	32	4.4	7	20.1
8	A2	165.7	43	3.9	12	13.8
9	K	117.0	32	3.7	16	7.3
10	O	146.4	33	4.4	11	13.3
11	B1	128.7	25	5.1	6	21.5
12	H	105.7	24	4.4	9	11.7
13	F	146.4	44	3.3	14	10.5
14	B2	113.6	25	4.5	12	9.5
15	E	71.3	19	3.8	5	14.3
16	J	90.3	24	3.8	10	9.0
17	D2	129.0	42	3.1	14	9.2
18	G	108.7	37	2.9	16	6.8
19	L	115.1	31	3.7	5	23.0
20	I	113.8	32	3.6	9	12.6
	Mean	121.5	32	4	10.5	11.6
	Total		626		209	

† Chromosome.
‡ Linkage group.
§ Chromosome length in centimorgans.
¶ Number of markers screened for each chromosome.
# Chromosome length per marker (total chromosome length / number of markers screened).
†† Number of polymorphic marker screened for each chromosome.
‡‡ Chromosome length per polymorphic marker (total chromosome length / number of polymorphic markers).

Table 4. Mean effect of SSR marker alleles on seed sucrose in 220 $F_{2:3}$ lines derived from MFS-553 x PI 243545 evaluated in Fayetteville, AR in 2010.

SSR marker	Chr.†	P1‡	P2§	Diff.¶	P > F	R^2
Satt242	9	6.59	6.28	-0.31	0.0246	0.03
Satt002	17	6.32	6.63	0.31	0.0415	0.03
Satt669	17	6.36	6.70	0.34	0.0280	0.03
Satt427	18	6.33	6.64	0.31	0.0259	0.03
Satt166	19	6.49	6.19	-0.30	0.0211	0.04
Sat_286	19	6.54	6.20	-0.34	0.0139	0.04

† Chromosome.
‡ MFS-553 allele effect.
§ PI 243545 allele effect.
¶ Allelic difference.

Table 5. Summary of SNP markers used in screening $F_{3:4}$ population derived from MFS-553 x PI 243545.

Chr.†	LG‡	Length(cM)§	No. SNP marker	No. SNP locus located	Average distance between SNP loci
1	D1a	169.3	28	21	8.1
2	D1b	112.8	19	18	6.3
3	N	275.0	83	62	4.4
4	C1	202.9	80	48	4.2
5	A1	219.6	47	26	8.4
6	C2	435.7	109	86	5.1
7	M	470.3	76	65	7.2
8	A2	371.0	106	76	4.9
9	K	344.9	165	112	3.1
10	O	576.8	60	40	14.4
11	B1	312.2	119	72	4.3
12	H	293.2	24	17	17.2
13	F	412.3	134	99	4.2
14	B2	227.5	159	68	3.3
15	E	396.3	132	100	4.0
16	J	534.9	237	130	4.1
17	D2	395.6	101	85	4.7
18	G	290.8	116	91	3.2
19	L	407.2	66	48	8.5
20	I	153.7	36	19	8.1
Average		330.1	95	64	5.1
Total		6602.0	1897	1283	

† Chromosome.
‡ Linkage group.
§ Chromosome length in centimorgans.

Table 6. Single marker analysis of variance for seed sucrose content in 94 F_3-derived lines derived from MFS-553 x PI 243545 evaluated in Fayetteville (FAY), AR in 2010, Argentina (ARG) in 2011, and Costa Rica (CR) in 2011.

Chr.†	SNP Marker	Position	2010	2011		Combined‡
		(cM)	FAY	ARG	CR	
			P-value	P-value	P-value	P-value
5	ss245668753	125.1	0.0048	0.0648	0.0051	0.0102
5	ss245672731	154.9	0.0047	0.1848	0.0960	0.0433
9	ss246796276	76.1	0.1213	0.0048	0.0007	0.0238
13	ss248014959	179.2	0.0202	0.0436	0.3131	0.0960
13	ss248007511	188.7	0.0121	0.0043	0.0615	0.0188
16	ss249186914	277.3	0.0472	0.0112	0.1792	0.0282
16	ss248953718	281.8	0.0241	0.0192	0.2233	0.0319
17	ss249412632	301.0	0.0357	0.0005	0.0426	0.0016
18	ss249519978	35.4	0.0678	0.2116	0.0607	0.0813
19	ss250098318	302.1	0.2063	0.0808	0.0562	0.1015

† Chromosome.
‡ Pooled data from Fayetteville in 2010, Argentina and Costa Rica in 2011.

Table 7. Mean effect of SNP marker alleles on seed sucrose content in 94 F_3–derived population derived from MFS-553 x PI 243545 in 2010 and 2011.

		2010			2011†			Combined‡			
Chr.	**/SNP**	**P1§**	**P2¶**	**Diff.#**	**P1**	**P2**	**Diff.**	**P1**	**P2**	**Diff.**	**R^2**
5	ss245672731	7.03	7.80	0.77	6.15	6.40	0.25	6.44	6.87	0.43	0.10
9	ss246796276	7.48	7.00	-0.48	6.51	5.95	-0.56	6.83	6.30	-0.53	0.11
13	ss248014959	7.20	7.77	0.57	6.14	6.44	0.30	6.49	6.88	0.39	0.06
13	ss248007511	7.10	7.87	0.77	6.04	6.55	0.51	6.39	6.99	0.60	0.13
16	ss249186914	6.24	7.33	1.09	5.45	6.18	0.73	5.73	6.56	0.83	0.08
16	ss248953718	7.06	7.34	0.28	5.80	6.20	0.40	6.21	6.58	0.37	0.10
17	ss249412632	7.13	7.74	0.61	6.06	6.52	0.46	6.42	6.92	0.50	0.12
18	ss249519978	7.06	7.58	0.52	6.07	6.37	0.30	6.40	6.77	0.37	0.08
19	ss250098318	7.43	7.15	-0.28	6.31	6.04	-0.27	6.69	6.41	-0.28	0.03

† Pooled data from Argentina and Costa Rica in 2011.
‡ Pooled data from Fayetteville in 2010, Argentina and Costa Rica in 2011.
§ MFS-553 allele effect.
¶ PI 243545 allele effect.
# Allelic difference.

Table 8. Comparison of physical map position of putative sucrose QTL predicted from $F_{2:3}$ population using SSR markers and from F_{3}-derived population using SNP markers.

Chr.†	QTL from $F_{2:3}$ population		QTL from F_3-derived population	
	Linked SSR marker	**Physical position (bp)‡**	**Linked SNP marker**	**Physical position (bp)**
5	Sat_344 – Sat_407	3691696 - 33718372	ss245668753	27447797
16	Sat_366 – Satt431	30036653 -35718642	ss249186914	32854900

† Chromosome.
‡ The position in soybean 'Williams 82' physical map based on the reference genome sequence data.

Table 9a. Average monthly temperatures (°C) in Upala, Costa Rica and Cordoba, Argentina.

Month	Dec	Jan	Feb	Mar	Apr
Costa Rica	25.2	25.1	25.3	26.4	27.3
Argentina	22.8	23.4	22.6	20.8	18.0

Information from http://www.world-climates.com.

Table 9b. Average monthly temperatures (°C) at Fayetteville, AR in 2010.

Month	Jun	July	Aug	Sep	Oct	Nov
Fayetteville	26.5	27.1	28.6	22.5	16.1	10.5

Information kindly provided by Dr. Kristofor Brye.

Table 10. Soil testing in Fayetteville, AR in 2012.

Soil sample	pH	EC	P	K	Ca	Mg	S	Na	Fe	Mn	Zn	Cu	B	LOI†
		umhos/cm	mg/Kg											%
Corner A	6.1	121	34.8	148	992	108	9.0	10.1	92	180	1.4	3.9	0.4	1.28
Corner B	6.3	130	38.2	146	1070	119	12.3	11.0	103	183	1.6	3.9	0.4	1.41
Corner C	6.1	167	37.9	104	823	94	8.1	9.3	102	202	1.4	4.4	0.3	1.01
Corner D	6.2	121	25.6	101	856	101	6.3	7.8	96	159	1.2	3.3	0.2	1.06

† loss-on-ignition.

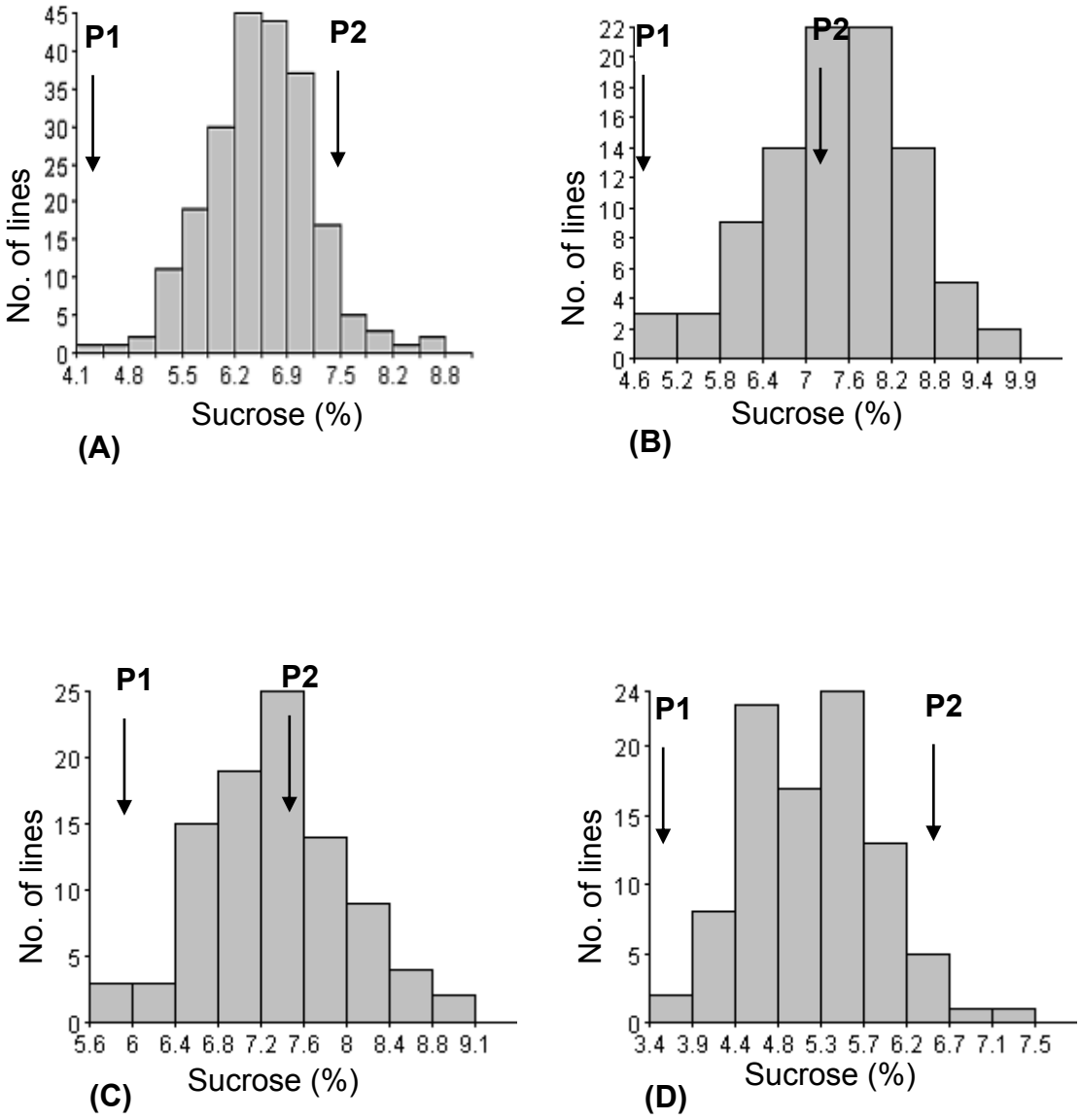


Figure 1. Frequency distribution of seed sucrose content in populations derived from MFS-553 (P1) x PI 243545 (P2) evaluated in four environments: (A) $F_{2:3}$ lines in Fayetteville, AR in 2010; (B) $F_{3:5}$ lines in Fayetteville, AR in 2011; (C) $F_{3:6}$ lines in Argentina in 2011; and (D) $F_{3:6}$ lines in Costa Rica in 2011.

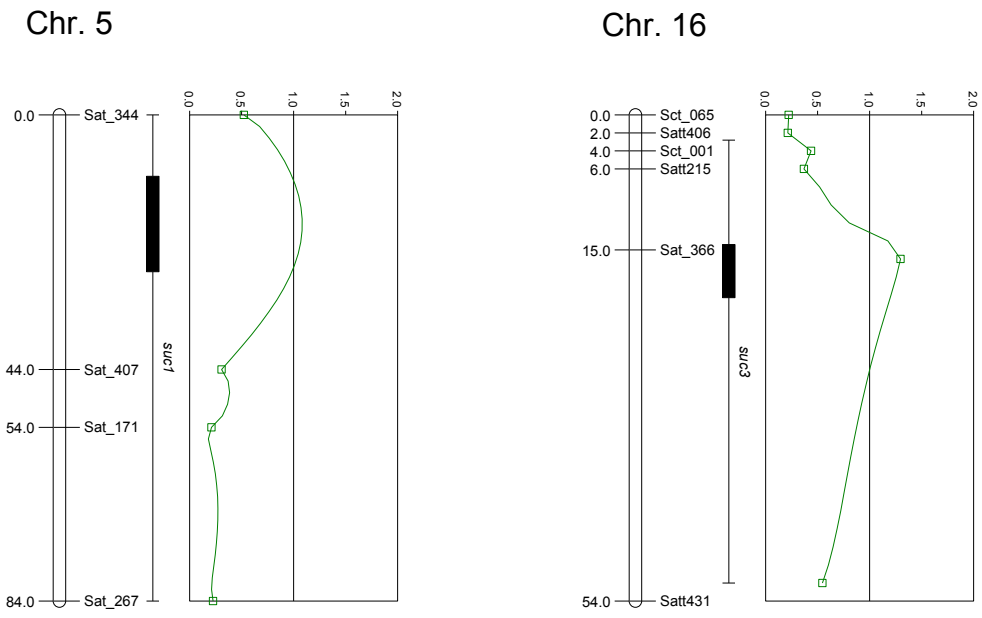


Figure 2. Composite interval mapping using SSR markers for seed sucrose QTL on chromosomes 5 (A) and 16 (B) in 220 $F_{2:3}$ lines from MFS-553 x PI 243545 evaluated in Fayetteville, AR in 2010.

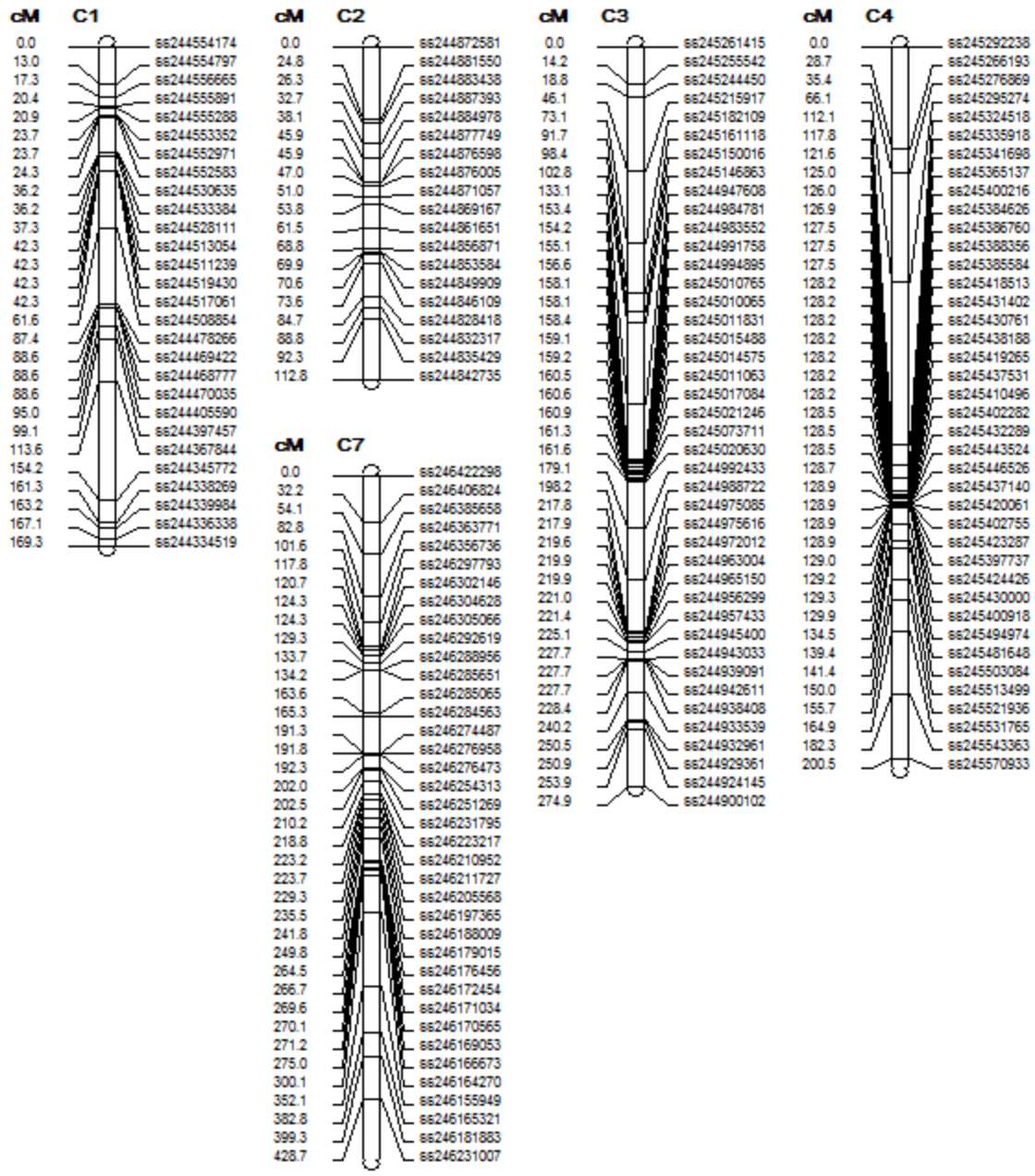


Figure 3A. A genetic map constructed for chromosomes 1, 2, 3, 4, and 7 using a $F_{3:4}$ mapping population derived from MFS-553 x PI 243545. A total of 1897 polymorphic SNP markers were mapped to 20 soybean chromosomes.

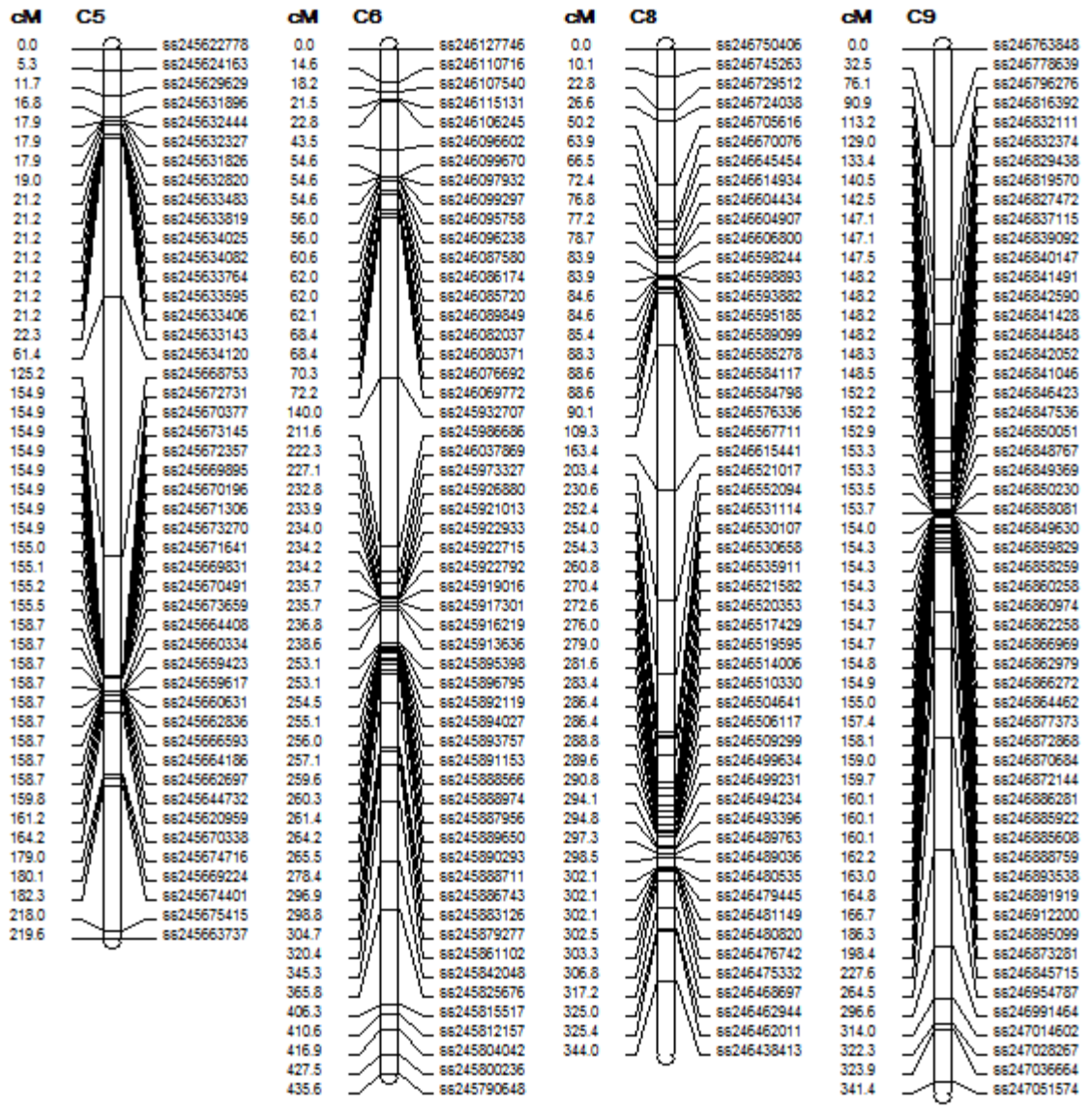


Figure 3B. A genetic map constructed for chromosomes 5, 6, 8, and 9 using a $F_{3:4}$ mapping population derived from MFS-553 x PI 243545. A total of 1897 polymorphic SNP markers were mapped to 20 soybean chromosome.

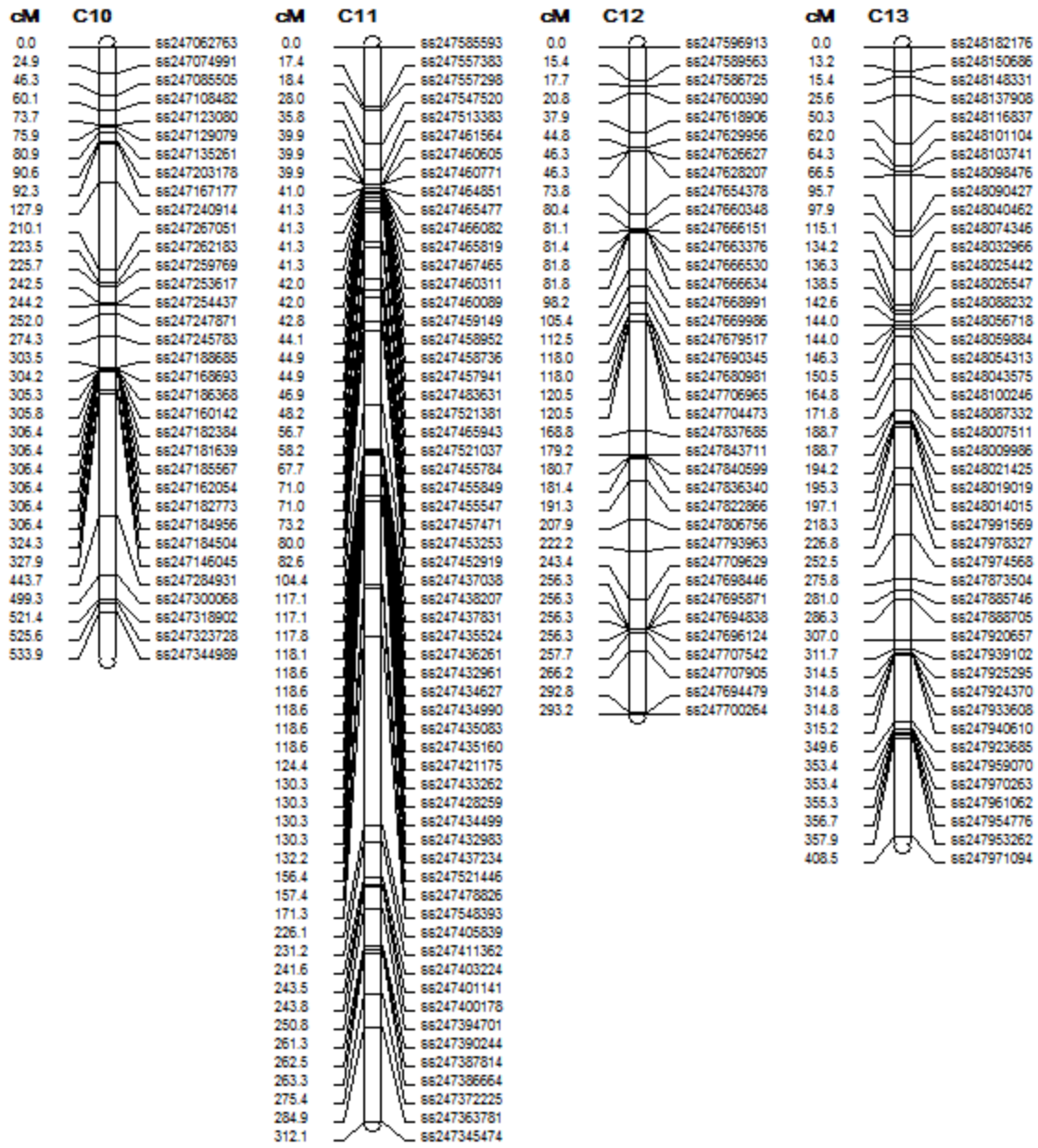


Figure 3C. A genetic map constructed for chromosomes 10, 11, 12, and 13 using a $F_{3:4}$ mapping population derived from MFS-553 x PI 243545. A total of 1897 polymorphic SNP markers were mapped to 20 soybean chromosome.

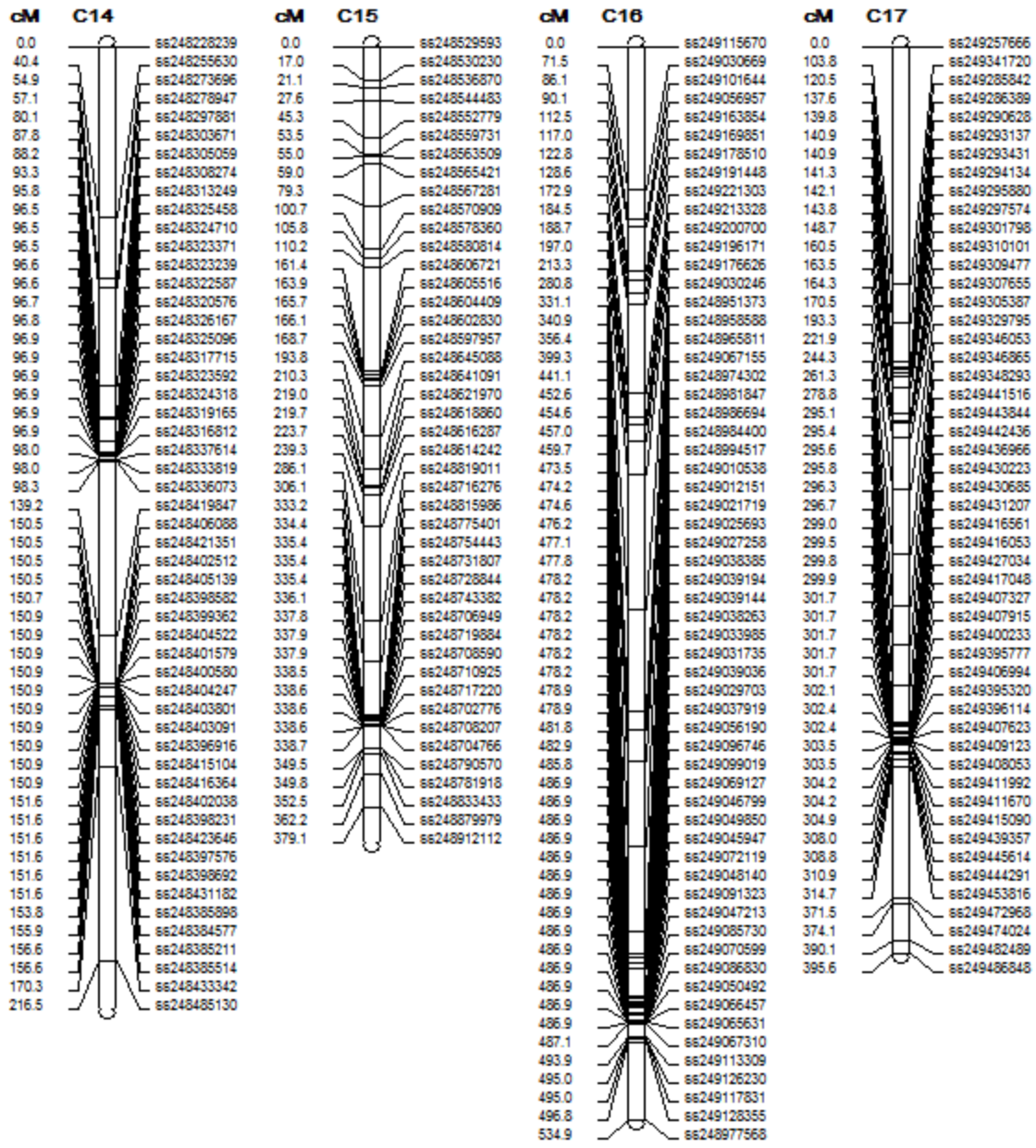


Figure 3D. A genetic map constructed for chromosomes 14, 15, 16, and 17 using a $F_{3:4}$ mapping population derived from MFS-553 x PI 243545. A total of 1897 polymorphic SNP markers were mapped to 20 soybean chromosomes.

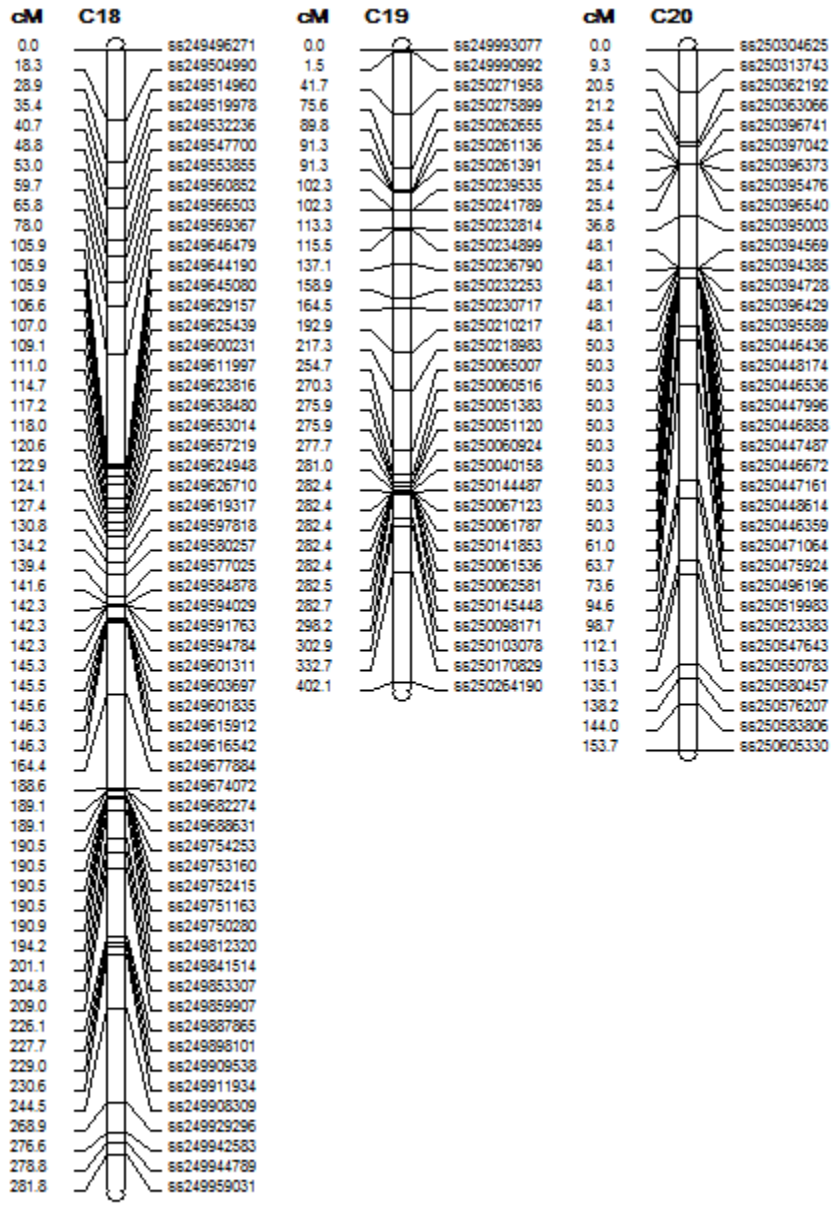


Figure 3E. A genetic map constructed for chromosomes 18, 19, and 20 using a $F_{3:4}$ mapping population derived from MFS-553 x PI 243545. A total of 1897 polymorphic SNP markers were mapped to 20 soybean chromosomes.

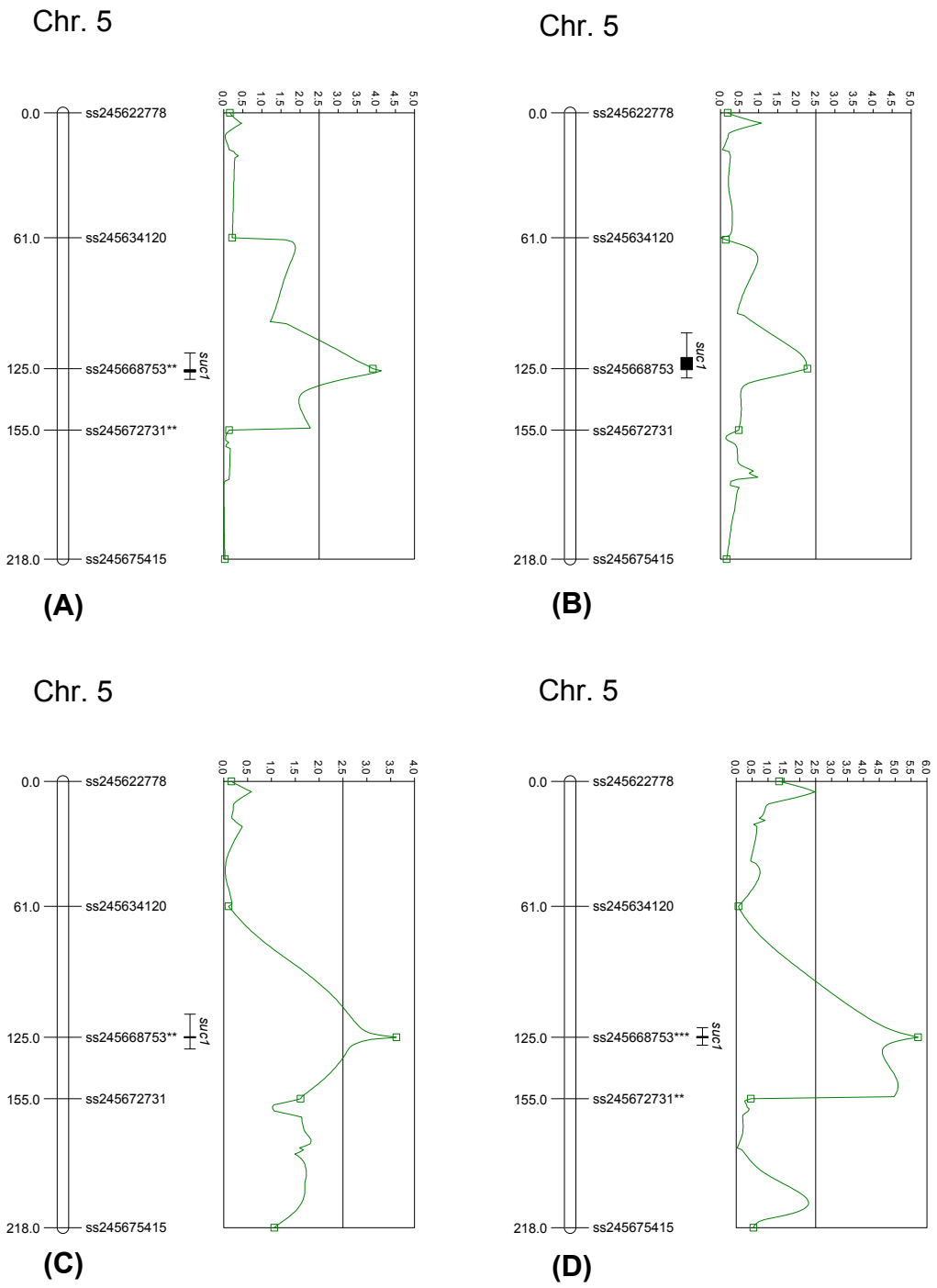


Figure 4. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 5 in 94 F_3-derived lines from MFS-553 x PI 243545 evaluated across three locations: (A) in Fayetteville, AR in 2010; (B) in Argentina in 2011; (C) in Costa Rica in 2011; (D) combined data across three locations and two years.

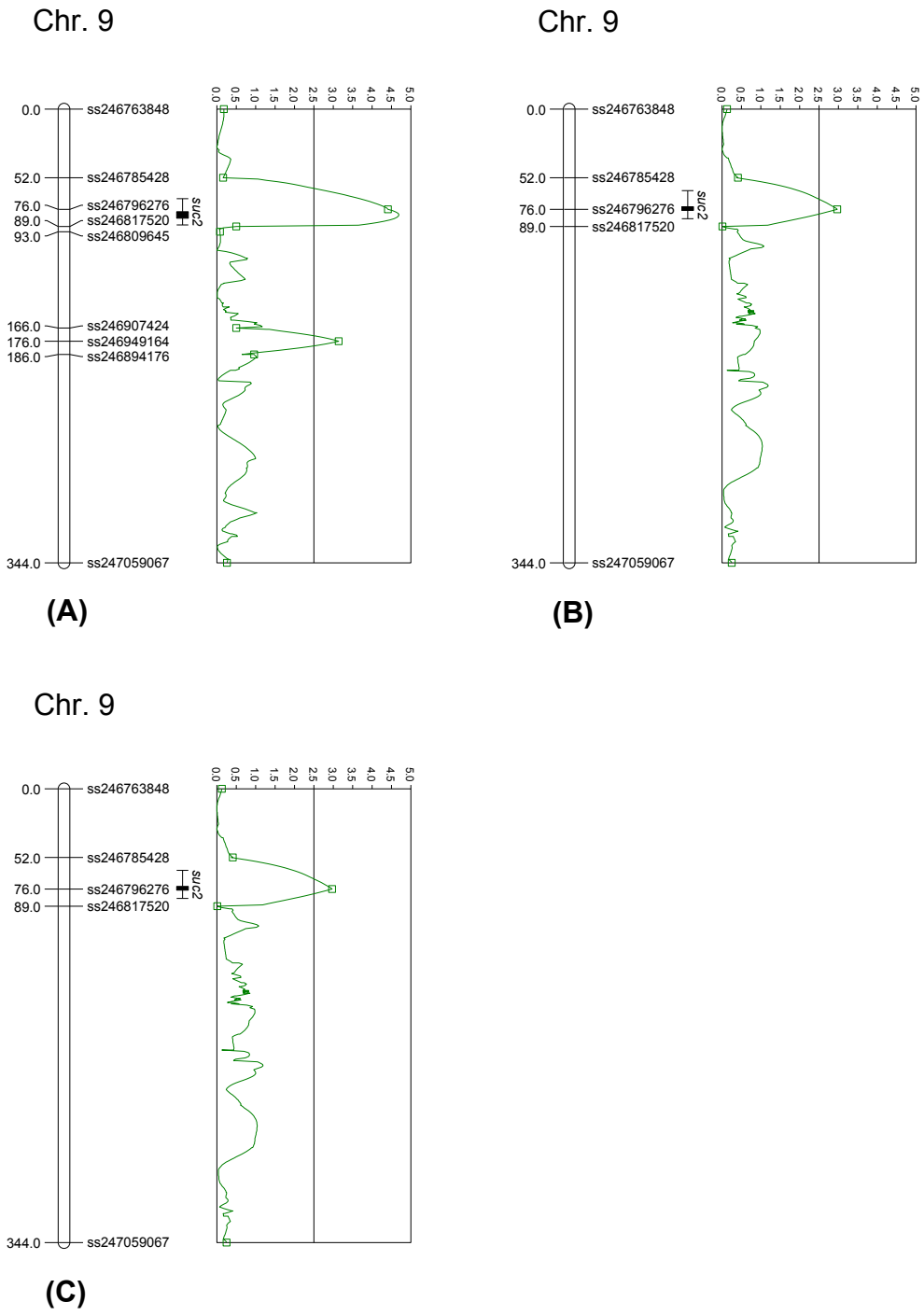


Figure 5. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 9 in 94 F_3-derived lines from MFS-553 x PI 243545 evaluated across two environments: (A) in Argentina in 2011; (B) in Costa Rica in 2011; (C) combine data across two locations in 2011.

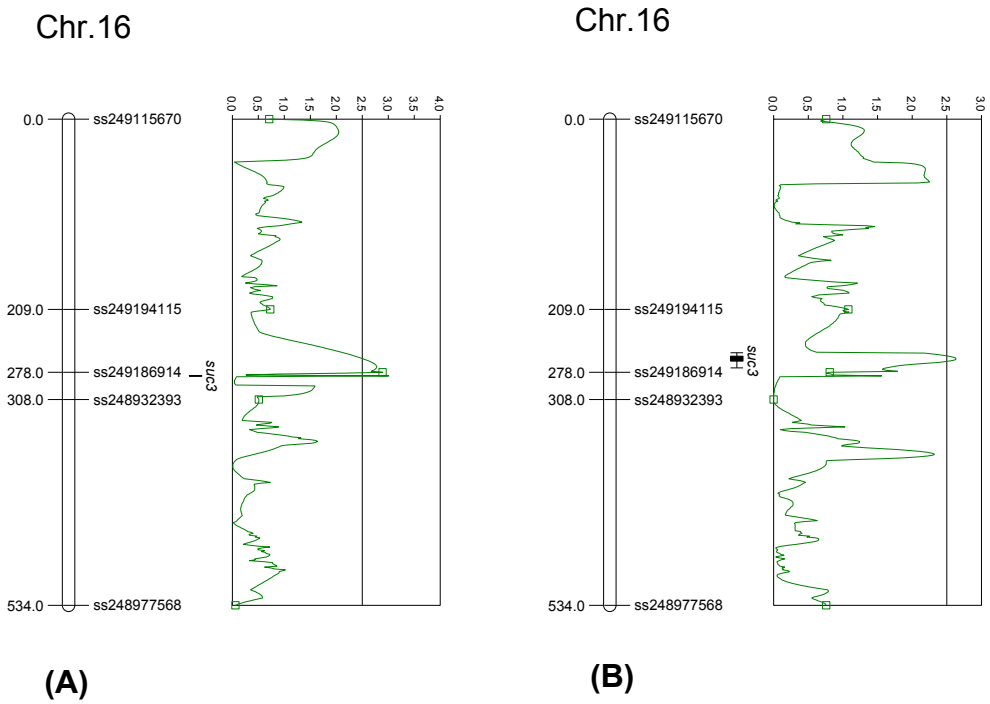


Figure 6. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 16 in 94 F_3-derived lines from MFS-553 x PI 243545 evaluated in two environments: (A) in Fayetteville in 2010; (B) in Argentina in 2011.

CHAPTER 3 IDENTIFICATION AND CONFIRMATION OF QUANTITATIVE TRAIT LOCI FOR STACHYOSE CONTENT IN SOYBEAN SEED

ABSTRACT

Stachyose is an unfavorable sugar in soybean meal causing flatulence for nonruminant animals. Understanding the genetic control of stachyose in soybean would facilitate the modification of stachyose content at the molecular level. The objectives of this study were to identify QTL associated with seed stachyose content using simple sequence repeat (SSR) markers and single nucleotide polymorphism (SNP) markers. A normal stachyose cultivar, Osage, was crossed with a low stachyose line, V99-5089, to develop a QTL mapping population. Two parents were screened with 33 SSR and 37 SNP markers randomly distributed on chromosome 10, and 1 SSR and 19 SNP markers surrounding a previously reported stachyose QTL region on chromosome 11. A total of 10 SSR and 18 SNP markers were polymorphic. Of these, 5 SSR and 16 SNP markers were used to screen the $F_{3:4}$ lines derived from Osage x V99-5089. Seed from $F_{3:5}$ and $F_{3:6}$ lines were analyzed for stachyose using high performance liquid chromatography. Following the linkage map construction, composite interval mapping and multiple interval mapping were conducted to locate putative QTL. Two stachyose QTLs were mapped to chromosome 10 and 11, explaining 11 and 81% of phenotypic variation for stachyose content, respectively. These stachyose QTLs were stable in distinct genetic backgrounds and across environments. The SSR / SNP markers linked with stachyose QTL could be used in breeding soybean lines with desired stachyose contents.

INTRODUCTION

Stachyose is a major soluble sugar followed by sucrose in soybean seed (Liu, 1997), accounting for 1.4 to 4.1% of the seed composition on a dry-weight basis (Hymowitz et al., 1972). Of the raffinose saccharides, stachyose is the most abundant and most nutritionally undesirable in the soybean (Cristofaro et al., 1974).

Stachyose in soybean meal causes flatulence and diarrhea for nonruminant animals (Kuriyama and Mendel, 1917; Wiggins, 1984; Hata et al., 1991), since nonruminant animals lack the α-(1, 6)-galactosidase enzyme to digest stachyose (Gitzelmann and Auricchio, 1965). Therefore, soybean meal supplied low metabolizable energy for nonruminant animals (Coon et al., 1990). In addition, the artificial removal of raffinose saccaride by ethanol extraction in soybean meal was reported to increase its digestibility from 52.1 to 63.3% (Leske et al., 1999). Therefore, the development of lines with low seed stachyose content would increase the metabolism rate of soybean meal and then improve the feed efficiency.

V99-5089 is a soybean line with low phytic acid, low stachyose and high sucrose content (Maroof et al. 2008). The stachyose content (0.36% - 0.60%) in V99-5089 was 10 times lower than in common soybean cultivars (Maroof et al., 2008). The stachyose content of progenies derived from different crosses between V99-5089 and other common soybean lines indicated that the low stachyose trait was heritable, suggesting that V99-5089 could be utilized to develop genetic populations by crossing with normal stachyose soybean lines and for characterizing the genetic mechanism of stachyose. The crosses between V99-5089 and other low stachyose soybean lines suggested that

V99-5089 had a distinct stachyose QTL or gene from other low stachyose sources previously reported by Sebastian et al. (2000) and Skoneczka et al. (2009).

The major stachyose QTL from V99-5089 was mapped in the interval between the MIPS1 gene (3.8 cM above Satt453) and Sat_331 on chromosome 11 and explained 28% of the phenotypic variance for seed stachyose content (Maroof et al., 2008). Satt453 was also applied to differentiate the low stachyose phenotype from the high stachyose type (Maroof et al., 2008). Another putative QTL for stachyose was mapped in the recombinant inbred line (RIL) population derived from the cross R95-1705 x MFL-552; this QTL was located between Satt262 and Sat_282 on chromosome 10 and explained 48% of the phenotypic variance for stachyose content (Jaureguy et al., 2009).

The objectives of this study were to confirm the pervious reported stachyose QTL on chromosome 10 and 11, further narrow down the stachyose QTL region using single nucleotide polymorphism (SNP) markers, and reveal the candidate stachyose genes. In this study, new QTLs/markers are supposed to be found or previously reported QTL may be confirmed in different genetic backgrounds.

MATERIALS AND METHODS

Parental materials

Osage is a maturity group V conventional cultivar, derived from ‘Hartz 5545’ x ‘KS4895’ and released by the Arkansas Agricultural Experimental Station (Chen et al. 2007). Osage has purple flowers, gray pubescence, and regular seed size with normal stachyose content (3.5%-5%), while V99-5089 was also a large-seeded maturity group

V line, which was derived from the cross between V71-370 and PI 87013 and released by Virginia Polytechnic and State University. V99-5089 has purple flowers, gray pubescence, and low stachyose content (<1%).

Population development and field experiment

The cross Osage x V99-5089 was made in 2007 at the Agricultural Experiment Station of the University of Arkansas, Fayetteville, AR. The F1 plants were grown and confirmed as true hybrids in the field in Fayetteville, AR in 2008. The F_2 population, consisting of approximately 1,200 plants, was grown in Fayetteville, AR in 2009 and 250 F2 plants were individually harvested to derive F2:3 lines. A sample of 60-100 seed from each $F_{2:3}$ line was sent to Costa Rica's winter nursery for generation advancement.

In the winter nursery, each $F_{2:3}$ line was grown in a 3-meter row with a 0.79-meter row spacing. From each $F_{2:3}$ row, a single plant (F_3) was randomly pulled and threshed to obtain recombinant inbred lines (RIL) for QTL mapping. Seed from the single F_3 plants pulled from the winter nursery were planted in single rows ($F_{3:4}$ lines) in Fayetteville, AR in summer 2010. $F_{3:5}$ seed from each row were bulk harvested in fall 2010. A total of 129 $F_{3:5}$ lines with adequate seed were grown in Costa Rica and Argentina. Plots were 3-meter long with 0.79-meter row spacing and were in a randomized complete block design with two replications. Seed from each plot were harvested separately for sugar analysis.

Soil analysis

Soybeans were planted in Fayetteville, AR in 2010, and were later sent to winter nurseries in Argentina and Costa Rica in 2011. The Fayetteville site is located in the

Ozark Highlands of northwest Arkansas at an elevation above sea level of 426 m (Mongabay, 2012). The winter nursery in Argentina is near Pergamino, which is located in eastern part of the country at an elevation of 62 m (Mongabay, 2012). At the Argentina site, the soil surface texture is silt loam with a pH near 5.5 and SOM concentration of approximately 2.9% (INTA, 2012). The winter nursery in Costa Rica is near Upala in the northwestern part of the country at an elevation of 46 m (Mongabay, 2012). At the Coast Rica site, the soil surface pH ranges from 6.0 to 6.4, while the soil texture and SOM concentration are unknown.

At all three sites used to grow soybean in this study, the soil was cultivated before planting. All soybean plots were planted as 3-m-long rows with a row spacing of 0.76 m. In addition, soybeans were irrigated during the growing season at all three sites.

Soil samples were collected from top 10 cm at the four corners (A, B, C, and D) of the study area in Fayetteville, oven-dried at 70^{o}C for 48 hr, and crushed and sieved to pass a 2-mm mesh screen. Sieved soil was used for soil particle-size analysis based on the 2-hr hydrometer method described by Arshad et al (1996). Sieved soil was also used to determine soil pH and electrical conductivity potentiometrically on a 1:2 (m/v) soil-to-water paste and Mehlich-3 extractable soil nutrients including P, K, Ca, Mg, S, Na, Fe, Mn, Zn, Cu and B (Watson et al., 1998; Donahue et al., 1983; Tucker, 1992). Soil organic matter was determined by loss-on-ignition (Combs and Nathan,1998; Wolf and Miller, 1998).

Sugar analysis

Sugar extraction followed the protocol previously described by Hou et al. (2009). Briefly, seed were ground to a fine powder using a coffee bean grinder (Krups®, Shelton, CT), then the powder was passed through a 100 µm standard testing sieve (VWR International, West Chester, PA) to obtain a homogeneous sample. Then, a 0.15 g sample was weighed and dissolved in 1.5 mL of distilled water, vortexed and then shaken horizontally at 200rpm for 20 min. followed by centrifugation at 10000 rpm for 10 min. Then, 500 µL of supernatant were transferred to a fresh tube. The sample was then dissolved with 700 µL acetonitrile for 30 min. followed by centrifugation at 10000 rpm for 10 min. Then, the supernatant was filtered through a 0.2-µm filter, and finally 24 µL of sugar extract were re-dissolved with 576 µl ddH_2O for high performance liquid chromatography (HPLC) analysis.

The HPLC system (Dionex DX500) consisted of a GS50 gradient pump, AS40 automated sampler, ED40 electrochemical detector, a chromeleon chromatography management system, and a CarboPac PA 10 pellicular anion-exchange resin column. A series of sugar concentrations of 10, 20, 40, 60, and 80µg/µL were used to construct the standard curve for glucose, fructose, sucrose, raffinose and stachyose. The sugar concentration was converted into mg/g based on dry seed matter.

DNA extraction

Tissue samples were bulk harvested from each row of 144 $F_{3:4}$ lines in summer 2010. Total genomic DNA was extracted using the CTAB (hexadecyltrimethyl ammonium bromide) method (Kisha et al., 1997). Briefly, tissue samples were ground in liquid nitrogen with a mortar and pestle and the extraction buffer was added to the

ground sample followed by chloroform:isoamyl alcohol (24:1) to remove the protein. The DNA was dissolved by 0.1 × TE buffer and the concentration was measured using Bio-Tek PowerWave XS Microplate Spectrophotometer (BioTek Instruments, Winooski, VT). The DNA solution was stored at -80°C.

Polymerase chain reaction and SSR marker screening

Polymerase chain reaction (PCR) was performed in an iCycler ThermalCycler (Bio-Rad, Hercules, CA). The total reaction volume for PCR was 15.2 µL, consisting of 4 µL template DNA (20ng/µl), 1µL primers (0.5 µM), 1 unit of Taq polymerase, 1.8 µL $MgCl_2$ (2.5 mM), 0.9 µL dNTP (2.5 mM), and 4.3 µL ddH_2O. PCR program was set as 94°C for 4 min (predenaturation) followed by 32 cycles of reaction, which consisted of 94°C for 25 s (denaturation), 47°C for 25 s (annealing), 68°C for 25 s (extension), and another cycle for final extension at 72°C for 7 min.

The PCR products were loaded on the 6% non-denaturing polyacrylamide gel electrophoresis (PAGE) (C.B.S. Scientific, San Diego, CA) and the gel was run at a constant 350V for 2 hours. The PCR products were stained with ethidium bromide and were visualized under the ultraviolet light (Wang et al., 2003).

SNP marker genotyping

A total of 129 $F_{3:4}$ lines were genotyped with SNP markers at the sequenom technology core, Washington University School of Medicine, St. Louis, MO. Two parents were screened with 33 SSR and 37 SNP markers (Song et al., 2010) which were randomly distributed on chromosome 10. In addition, one SSR and 19 SNP markers surrounding a previous reported stachyose QTL region on chromosome 11 were used

to screen the parents. Markers that were polymorphic between parental lines were used to screen the $F_{3:4}$ lines.

Data analysis

The Shapiro-Wilk (*w*) statistic of the JMP 9.0 (SAS Institute Inc.,Cary, NC) was used to test the normality of stachyose content distribution for F_3-derived lines. Chi-square test for goodness-of-fit to two-gene model was calculated using the equation (1).

$$X^2 = \Sigma [(\text{Observed} - \text{Expected})^2 / \text{Expected}] \quad (1)$$

The associations between seed stachyose and molecular markers were determined by single factor analysis of variance (ANOVA) analysis at the 0.05 significance level using the PROC GLM procedure of SAS 9.0. A linkage map was constructed using JoinMap 4.0 (Van Ooijen, 2006) and the threshold for logarithm of odds (LOD) for linkage groups construction was set as 3.0. Regression mapping of each linkage group was performed with Haldane mapping function (Haldane, 1919). The composite interval mapping was performed using Qgene (Joehanes et al., 2008). One thousand permutation with a walk speed of 1cM and experiment-wise α=0.05 was adopted to establish the empirical significance threshold (Churchill and Doerge, 1994). Multiple-interval mapping analysis (MIM) was performed to determine the optimum position of QTL and the interaction among QTL, MIM was performed based on the model $c(n) = \ln(n)$ with a walk speed of 1 cM. MapChart (Voorrips, 2002) was used to create the charts of linkage maps and QTLs by virtue of the map files exported by JoinMap 4.0 and QTL Cartographer 2.5.

RESULTS

Soil testing results

Soil testing results showed that the soil at the Fayetteville site is a Captina silt loam (fine-silty, siliceous, active, mesic Typic Fragiudult; SSS-NRCS-USDA, 2012) with 33.2 % sand, 50.3% silt, and 16.5% clay in the top 10 cm and other soil chemical properties including soil pH ranging from 6.1 to 6.3 with an average of 6.2 and soil organic matter (SOM) concentration ranging from 1.0 to 1.4% with an average of 1.2% (Table 6).

Phenotypic data

Osage showed a normal stachyose and V99-5089 consistently exhibited a very low stachyose across locations and years (Table 1). The normality tests showed that he F3-derived population did not follow a normal distribution in all three locations (Table 1), indicating the stachyose is likely a qualitative trait controlled by a few major genes. When the data were plotted for population distributions, seed stachyose of 129 F_3-derived lines showed an obvious bimodal distribution for all three locations, suggesting that stachyose content is controlled by one or two major genes (Fig.1). When the phenotypic data were tested using Chi-square tests for goodness-of-fit to the single gene or two-gene model, a satisfactory fit to a ratio of 15 normal :1 low stachyose was obtained, indicating a presence of two genes segregating in the population (Table 2).

Linkage map construction

Of the 56 SNP markers screened, 18 markers were polymorphic between the parental lines, and 10 out of 34 SSR markers screened also showed polymorphism between the parents. Among the polymorphic markers, sixteen SNP (Table 3) and five

SSR markers were used to construct the linkage map. The grouping and position of most markers screened were consistent with the public soybean linkage map (Song et al., 2010).

Single marker analysis

Seed stachyose QTLs were analyzed separately and jointly using data from three locations in two years. Thirteen markers on chromosome 10 were significant ($P < 0.05$) and environmentally stable, except that Satt259 and Satt173 were not significant for Costa Rica in 2011 (Table 4). All six markers on chromosome 11 were also and significant environmentally stable (Table 4). These results indicate that the two major genes for stachyose reside on chromosome 10 and 11, respectively.

Allelic effect

The homozygous lines with the Osage allele had a greater stachyose content than those with the V99-5089 allele at all 19 marker alleles on chromosome 10 and 11 (Table 5). The stachyose content differences were from 1.14 to 2.42% between parental SNP marker alleles on chromosome 11 and from 0.58 to 0.98% at marker alleles on chromosome 10 (Table 5). Single nucleotide polymorphism markers BARC-018869-03031 and BARC-013547-01157 on chromosome 11 independently explained 59% and 52% of the phenotypic variance, while BARC-018557-03202, BARC-044669-08757, ss247579175 and Satt453 explained 28%, 28%, 29%, and 19% of phenotypic variation, respectively (Table 5). On chromosome 10, markers explained from 6 to 14% of the phenotypic variation for seed stachyose content (Table 5). These data suggested that

chromosome 11 carry a major gene/QTL while chromosome 10 contains a minor gene/QTL for stachyose content.

QTL analysis

Composite interval mapping analysis revealed a stachyose QTL on chromosome 10 that was located on the interval between Sat_282 and BARC-029531-06209 and consistent across three location in two years (Fig. 3). This QTL was designated as *Sta1*. For the combine data, *Sta1* had a LOD score of 3.1 and explained 11% of phenotypic variation for seed stachyose content (Fig. 3).

A strong QTL on chromosome 11 was identified in the interval between BARC-013547-01157 and BARC-018869-03031 and it was consistent across three locations in two years. This QTL was designated as *Sta 2*, with a LOD score of 47, and explained 81% of phenotypic variation for seed stachyose content for the combine data (Fig. 4D).

Multiple interval mapping analysis showed that no significant epstasis interaction existed between *sta1* and *sta2*. The optimum positions indicated by the MIM analysis were 40.3 cM on Chromosome 10 for *sta1* and 6.0 cM on chromosome 11 for *sta2*, respectively. These results indicated that the two genomic regions on chromosome 10 and 11 are likely the genetic loci for the two stachyose genes and that these two genes are independent of each other.

DISCUSSION

The average seed stachyose contents of both parents and their progenies were consistent among different locations, suggesting that the stachyose content is stable

across the environments. These results are in agreement with those from other studies indicating that stachyose was genotype dependent (Wolf et al., 1982; Kumar et al., 2010). These results also suggest that stachyose is a qualitative trait which is controlled by single genes and is not largely influenced by the environmental conditions.

In this study, a major stachyose QTL was found on chromosome 11. This QTL was located in the same region for a QTL reported previously by Maroof et al. (2008). This result was not surprising given that V99-5089 was a common parent used in both studies. Therefore, we conclude that a major QTL for stachyose has been confirmed in different genetic backgrounds and across different environments. This QTL appeared to act as a single gene with very large effect (81%) on stahyose with little or no environmental impact. The presence of this QTL/gene alone would change stachyose content by up 2.4%. Therefore, any makers associated with this QTL/gene would be very useful in marker-assisted breeding for low stachyose content.

In this study, we also found a minor QTL for stachyose on chromosome 10. This QTL resides in a similar region for a stachyose QTL identified in a different parent (Jaureguy et al. 2009). The stachyose QTL identified in my study is in a narrower region. It appears that this QTL region identified in two independent studies contain the same QTL/gene for stachyose. Therefore, we claim that a new minor stachyose QTL is identified in V99-5089. This QTL was not discovered in the original study with V99-5089 (Maroof et al. 2008).

In our study, we used two different marker systems to detect QTL for stachyose. Both SSR and SNP revealed the same two QTL for stachyose. In addition, the QTL

analysis matches perfectly with the Chi-square analysis for two independent genes for stachyose. Therefore, we conclude that stachyose is conditioned by two separate genes, one on chromosome 11 with large effect and the other on chromosome 10 with small effect. These two genes are completely independent with no epistatic effect. This conclusion is further supported by the allele effect estimates in the QTL analysis where the QTL/gene on chromosome 11 contributes the decrease of stchyose by up to 2.42% and the other minor QTL on chromosome 10 contributes up to 1% stachyose reduction. The combination of the two QTL/genes would theoretically change stachyose content by a total of 3.42% considering the additive effect with no epistasis of the two genes. The two gene effect on stachyose appears to be close to the difference (3.46%) in stachyose between the two parents. This analysis further supports the two-independent-gene conclusion. Therefore, the SNP and SSR markers identified in his study would be very effective for marker-assisted selection for low stachyose content.

REFERENCES

Arshad, M.A., B. Lowery, and B. Grossman. 1996. Physical tests for monitoring soil quality. p. 123-141. *In* J.W. Doran and A.J. Jones (ed.) Methods for assessing soil quality. SSSA Spec. Publ. 49. SSSA, Madison, WI.

Chen, P., C.H. Sneller, L.A. Mozzoni and J.C. Rupe. 2007. Registration of 'Osage' soybean. J.P.R. 1: 89-92.

Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963-971.

Combs, S.M. and M.V. Nathan. 1998. Recommended chemical soil test procedures for the N central region. Soil organic matter. p. 53-59. MO Agricultural Experiment Station SB 1001. Columbia, MO.

Coon, C.N., K.L. Leske, O. Akavanichan, and T.K. Cheng. 1990. Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. Poultry Sci. 69:787-793.

Cristofaro, E., F. Mottu, and J.J. Wuharmann. 1974. Involvement of the raffinose family of oligosaccharides in flatulence. Pp 313-335. In: H. L. Sipple and K. W. McNutt (eds.) Sugars in Nutrition. Academic Press. London, UK.

Donahue, S.J. 1983. Determination of soil water pH. Reference soil test methods for the S region of the US. p. 1-4. S Coop. SB. 289. U of G College of Agriculture Experiment Station. Athens, GA.

Donahue, S.J. 1983. Determination of specific conductance in supernatant 1:2 soil:water solution. Reference soil test methods for the S region of the US. p. 8-14. S Coop. SB. 289. U of G College of Agriculture Experiment Station. Athens, GA.

Gitzelmann, R., and S. Auricchio. 1965. The handling of soy alpha-galactosidase by a normal and galactosemic child. Pediatrics 36:231-238.

Haldane, J.B.S. 1919. The combination of linkage values, and the calculation of distance between the loci of linked factors. J. Genet. 8:299-309.

Hata, Y., M.Yamamoto, and K. Nakajima. 1991. Effects of soybean oligosaccharides on human digestive organs: Estimation of fifty percent effective dose and maximum non-effective dose based on diarrhea. J. Clin. Biochem. Nutr. 10:135-144.

Hou, A., P. Chen, J. Alloatti, D. Li, L. Mozzoni, B. Zhang, and A. Shi. 2009. Genetic variability of seed sugar content in worldwide soybean germplasm collections. Crop Sci. 49:903-912.

Hymowitz, T., and F.I. Collins. 1974. Variability of sugar content in seed of *Glycine max* (L.) Merrill and *G. soja* Sieb and Zucc. Agron. J. 66: 239-240.

Leske, K.L., and C.N. Coon. 1999. Nutrient content and protein and energy digestibilities of ethanol extracted, low α-galactoside soybean meal as compared to intact soybean meal. Poult Sci. 78:1177-1183.

Jaureguy, L.M. 2009. Identification of molecular markers associated with seed size, protein and sugar content in soybean. M.S.Thesis, University of Arkansas, Fayetteville, AR.

JMP, Version 9. 2010. SAS Institute Inc.,Cary, NC.

Joehanes, R., and J.C. Nelson. 2008. QGene 4.0, an extensible Java QTL-analysis platform. Bioinformatics 24:2788-2789

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317–1325.

Kumar, V., A. Rani., L. Goyalt., A.K. Dixit., J.G. Manjaya., J. Dev., and M. Swamy. 2010. Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. J. Agric. Food Chem. 58:5081-5085.

Kuriyama, S., and L.B. Mendel. 1917. The physiological behavior of raffinose. J. Biol. Chem. 31:125-147.

Liu, K. 1997. Soybeans: Chemistry, Technology, and Utilization. p.532. Chapman & Hall, NY.

Maroof, M.A.S., and G.R. Buss. 2008. Low phytic acid, low stachyose, high sucrose soybean lines. U.S. Patent 12/033,830, field Feb.19, 2008.

Maupin, L.M., M.L. Rosso, and K.M. Rainey. 2011. Environmental effects on soybean with modified phosphorus and sugar composition. Crop Sci. 51:642-650.

Mongabay. 2012. http://population.mongabay.com/ (accessed 17July 2012).

National Institute of Agricultural Technology (INTA), Argentina. 2012. http://anterior.inta.gov.ar/suelos/cartas/series/Pergamino.htm (accessed 17July 2012).

Sebastian, S.A., P.S. Kerr, R.W. Pearisten, and W.D. Hitz. 2000. Soybean germplasm with novel genes for improved digestibility. p. 56-74. J.K. Drackely (ed.). In: Soy in animal nutrition. Federation of Animal Sci. Soc., Savoy, IL.

Skoneczka, J.A., M.A. Saghai Maroof, C.Shang, and G. R. Buss. 2009. Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI00508. Crop Sci. 49:247-255.

Soil Survey Staff (SSS), Natural Resources Conservation Service (NRCS), United States Department of Agriculture (USDA). 2012. Official Soil Series Descriptions. Available online at http://soils.usda.gov/technical/classification/osd/index.html. (accessed 18 July, 2012).

Song, Q.J., G.F. Jia, Y.L. Zhu, D. Grant, R.T. Nelson, E.Y. Hwang, D.L. Hyten, and P.B. Cregan. 2010. Abundance of SSR motifs and development of candidate polymorphic SSR Markers (BARCSOYSSR_1.0) in soybean. Crop Sci. 50:1950-1960.

Tucker, M.R. 1992. Determination of phosphorous by Mehlich 3 extraction. p. 6-8. In: S.J. Donohue (ed.), Soil and media diagnostic procedures for the Southern Region of the United States. Vir. Agric. Exp. Stn. Ser. Bull. 374. Blacksburg, VA.

Van Ooijen, J.W. 2006. JoinMap 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, the Netherlands.

Voorrips, R.E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77-78.

Wang, D., J. Shi, S.R. Carlson, P.B. Cregan, R.W. Ward, and B.W. Diers. 2003. A low-cost, high throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA marker. Crop Sci. 43:1828-1832.

Watson, M.E. and J.R. Brown. 1998. pH and lime requirements. N Central Regional Research Pub. 221:13-16.

Wiggins, H.S. 1984. Nutritional value of sugars and related compounds undigested in the small gut. Proc. Nutr. Soc. 43:69-75.

Wolf, R.B., J.F. Cavins., R. Kleiman., and L.T. Black. 1982. Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty acids, amino acids and sugars. J. Am. Oil Chem. Soc. 59:230-232.

Wolf, A.M. and R.O. Miller. 1998. Development of a NAPT program for soil and plant analysis. Commun. Soil Sci. Plant Anal. 29:1685-1690.

Table 1a. Seed stachyose content (%) using 129 F_3-derived lines from Osage x V99-5089 evaluated across three locations and over two years.

	Mean	SD†	Range	Prob< W‡
2010 Fayetteville $F_{3:5}$	3.30	1.12	0.19 - 5.08	< 0.0001
2011 Argentina $F_{3:6}$	3.72	1.23	0.23 - 5.38	< 0.0001
2011 Costa Rica $F_{3:6}$	3.19	1.02	0.20 - 4.77	< 0.0001

† Standard deviation.
‡ Sharpiro-Wilk value.

Table 1b. Seed stachyose content of Osage and V99-5089 evaluated across three locations and over two years.

	Osage		V99-5089	
	Mean	**SD†**	**Mean**	**SD†**
2010 Fayetteville	4.96	0.30	0.43	0.03
2011 Argentina	4.46	0.24	0.28	0.05
2011 Costa Rica	3.13	1.00	0.31	0.06

† Standard deviation.

Table 2. Chi-square test for goodness-of-fit to a two-gene model for 2010, 2011 and two years combination.

	Range	No. normal stachyose lines	No. low stachyose lines	X^2 (15:1)	P-Value
Fayetteville, 2010					
Osage	4.53-5.18	118	11	1.11	0.29
V99-5089	0.40-0.46				
$F_{3:5}$ lines	0.19-5.08				
Argentina, 2011					
Osage	4.12-4.76	121	8	0.001	0.97
V99-5089	0.21-0.37				
$F_{3:6}$ lines	0.23-5.38				
Costa Rica, 2011					
Osage	1.98-4.44	126	3	3.30	0.07
V99-5089	0.23-0.39				
$F_{3:6}$ lines	0.20-4.77				
Combine					
Osage	1.98-5.18	126	3	3.30	0.07
V99-5089	0.21-0.46				
F_3-derived lines	0.22-5.05				

Table 3. Summary of SNP markers used in screening the F_3-derived population from Osage x V99-5089.

Chr.†	LG‡	Length(cM)	No. SNP marker mapped	No. SNP locus located separately	Average distance between SNP loci (cM)
10	O	111.8	13	11	10.2
11	B1	23.4	5	5	4.7
Average		66	9	8	7.1
Total		132	18	16	

† Chromosome.
‡ Linkage group.

Table 4. Single marker analysis of variance for seed stachyose of 129 F_3 – derived population from the cross Osage x V99-5089 evaluated at Fayetteville (FAY), AR in 2010, Argentina (ARG) in 2011, and Costa Rica (CR) in 2011.

Chr.†	SSR/SNP Marker	Position (cM)	2010	2011		Combined‡
			FAY	**ARG**	**CR**	
			P-value	**P-value**	**P-value**	**P-value**
10	Satt259	0.0	0.0159	0.0040	0.0625	0.0150
10	BARC-017045-02182	15.6	0.0006	0.0003	0.0110	0.0011
10	BARC-019105-03305	23.3	0.0005	0.0002	0.0042	0.0006
10	Satt345	23.6	0.0015	0.0009	0.0128	0.0022
10	BARC-040251-07694	24.7	0.0002	0.0003	0.0030	0.0005
10	Satt173	31.9	0.0017	0.0126	0.0953	0.0148
10	BARC-044745-08785	36.9	0.0002	0.0004	0.0201	0.0012
10	Sat_282	40.2	< 0.0001	< 0.0001	0.0020	< 0.0001
10	BARC-029531-06209	42.1	0.0001	0.0002	0.0091	0.0006
10	BARC-016773-02317	44.7	0.0009	0.0013	0.0110	0.0023

10	ss247264293	59.6	0.0071	0.0025	0.0073	0.0045
10	ss247264289	59.6	0.0189	0.0064	0.0182	0.0122
10	ss247264305	59.6	0.0079	0.0026	0.0068	0.0048
11	BARC-013547-01157	0	< 0.0001	< 0.0001	< 0.0001	< 0.0001
11	BARC-018869-03031	10.8	< 0.0001	< 0.0001	< 0.0001	< 0.0001
11	BARC-044669-08757	20.2	< 0.0001	< 0.0001	< 0.0001	< 0.0001
11	ss247579175	20.2	< 0.0001	< 0.0001	< 0.0001	< 0.0001
11	BARC-018557-03202	20.2	< 0.0001	< 0.0001	< 0.0001	< 0.0001
11	Satt453	23.4	< 0.0001	< 0.0001	< 0.0001	< 0.0001

† Chromosome.
‡ Pooled data from Fayetteville in 2010, Argentina and Costa Rica in 2011.

Table 5. Mean effect of SSR/SNP marker alleles on seed stachyose content in 129 F_3-derived population from Osagex V99-5089 in 2010 and 2011.

		2010			2011†			Combined‡			
Chr.§	**SSR/SNP**	P3¶	P4#	Diff.††	P3	P4	Diff.	P3	P4	Diff.	R^2
10	Satt259	3.52	2.87	0.65	3.67	3.03	0.64	3.62	2.98	0.64	0.06
10	BARC-017045-02182	3.61	2.81	0.80	3.64	3.02	0.62	3.62	2.95	0.67	0.11
10	BARC-019105-03305	3.61	2.81	0.80	3.69	2.99	0.70	3.66	2.93	0.73	0.12
10	Satt345	3.60	2.83	0.77	3.68	3.00	0.68	3.65	2.95	0.70	0.09
10	BARC-040251-07694	3.62	2.78	0.84	3.68	2.98	0.70	3.66	2.91	0.75	0.12
10	Satt173	3.65	2.81	0.84	3.68	3.10	0.58	3.67	2.99	0.68	0.07
10	BARC-044745-08785	3.65	2.76	0.89	3.69	3.01	0.68	3.68	2.93	0.75	0.11
10	Sat_282	3.68	2.70	0.98	3.74	2.91	0.83	3.72	2.84	0.88	0.14
10	BARC-029531-06209	3.67	2.76	0.91	3.73	2.99	0.74	3.70	2.91	0.79	0.11
10	BARC-016773-02317	3.61	2.77	0.84	3.66	2.98	0.68	3.64	2.91	0.73	0.10
10	ss247264293	3.55	2.89	0.66	3.72	3.02	0.70	3.65	2.98	0.67	0.08

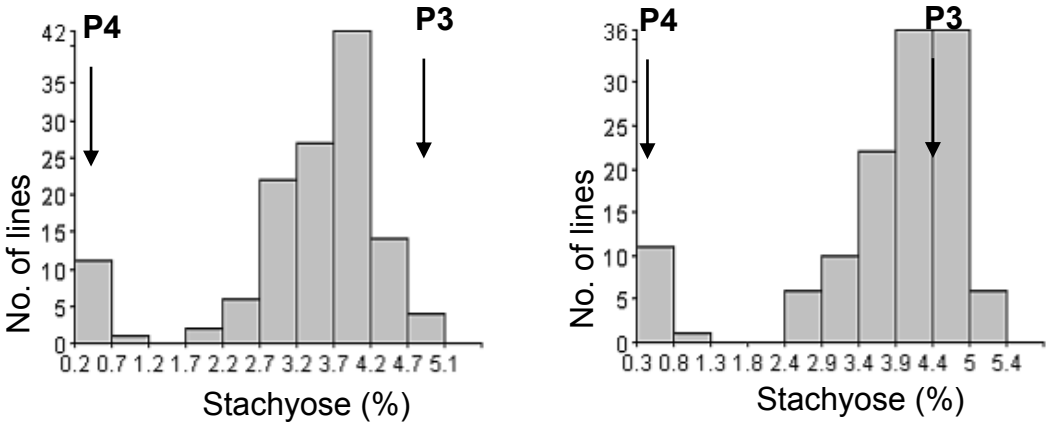
10	ss247264289	3.50	2.89	0.61	3.67	3.02	0.65	3.61	2.98	0.63	0.07
10	ss247264305	3.55	2.89	0.66	3.71	3.02	0.69	3.65	2.98	0.67	0.08
11	BARC-013547-01157	3.81	1.52	2.29	3.96	1.58	2.38	3.90	1.56	2.34	0.52
11	BARC-018869-03031	3.90	1.48	2.42	4.02	1.67	2.35	3.98	1.60	2.38	0.59
11	BARC-018557-03202	3.81	2.31	1.50	3.94	2.54	1.40	3.90	2.46	1.44	0.28
11	BARC-044669-08757	3.81	2.31	1.50	3.94	2.54	1.40	3.90	2.46	1.44	0.28
11	ss247579175	3.81	2.31	1.50	3.94	2.54	1.40	3.90	2.46	1.44	0.29
11	Satt453	3.80	2.55	1.25	3.93	2.79	1.14	3.88	2.71	1.17	0.19

† Pooled data from Argentina and Costa Rica in 2011.
‡ Pooled data from Fayetteville in 2010, Argentina and Costa Rica in 2011.
§ Chromosome.
¶ Osage allele effect.
# V99-5089 allele effect.
†† Allelic difference.

Table 6. Soil testing in Fayetteville, AR in 2012.

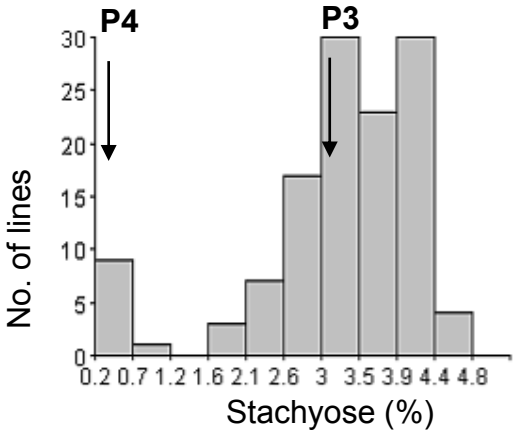
Soil sample	pH	EC	P	K	Ca	Mg	S	Na	Fe	Mn	Zn	Cu	B	LOI†
		umhos/cm	mg/Kg											%
Corner A	6.1	121	34.8	148	992	108	9.0	10.1	92	180	1.4	3.9	0.4	1.28
Corner B	6.3	130	38.2	146	1070	119	12.3	11.0	103	183	1.6	3.9	0.4	1.41
Corner C	6.1	167	37.9	104	823	94	8.1	9.3	102	202	1.4	4.4	0.3	1.01
Corner D	6.2	121	25.6	101	856	101	6.3	7.8	96	159	1.2	3.3	0.2	1.06

† loss-on-ignition.



.

(A) **(B)**



(C)

Figure 1. Frequency distribution of seed stachyose content in populations derived from Osage (P3) x V99-5089 (P4), $F_{3:5}$ lines at Fayetteville, 2010 (A), $F_{3:6}$ lines in Argentina, 2011(B), and $F_{3:6}$ lines in Costa Rica, 2011(C).

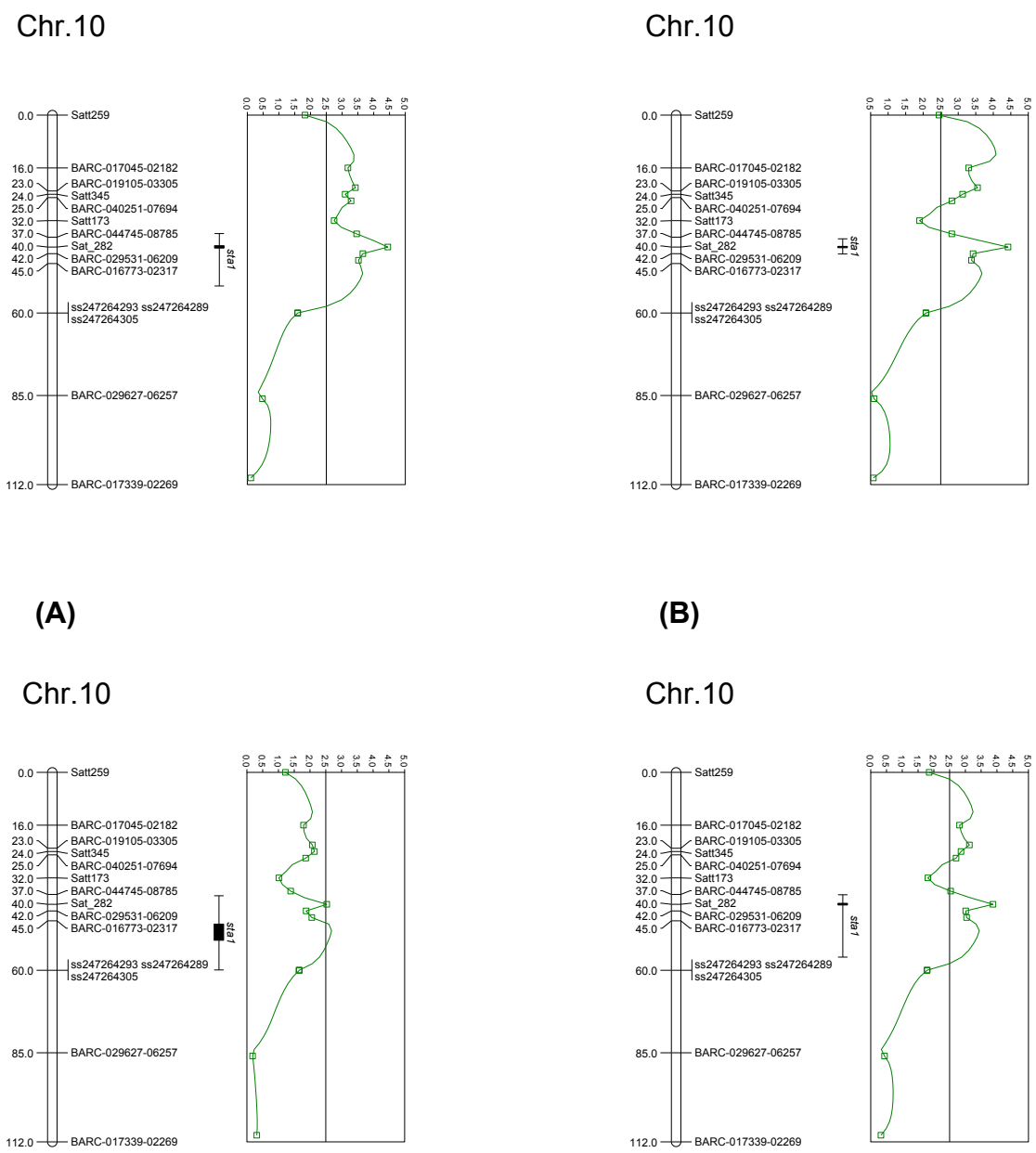


Figure 2. Composite interval mapping for seed stachyose QTL on chromosome 10 in 129 F_3-derived lines from Osage x V99-5089 evaluated across three environments: (A) in Fayetteville in 2010; (B) in Argentina in 2011; (C) in Costa Rica in 2011; (D) combine data across 3 locations in 2 years.

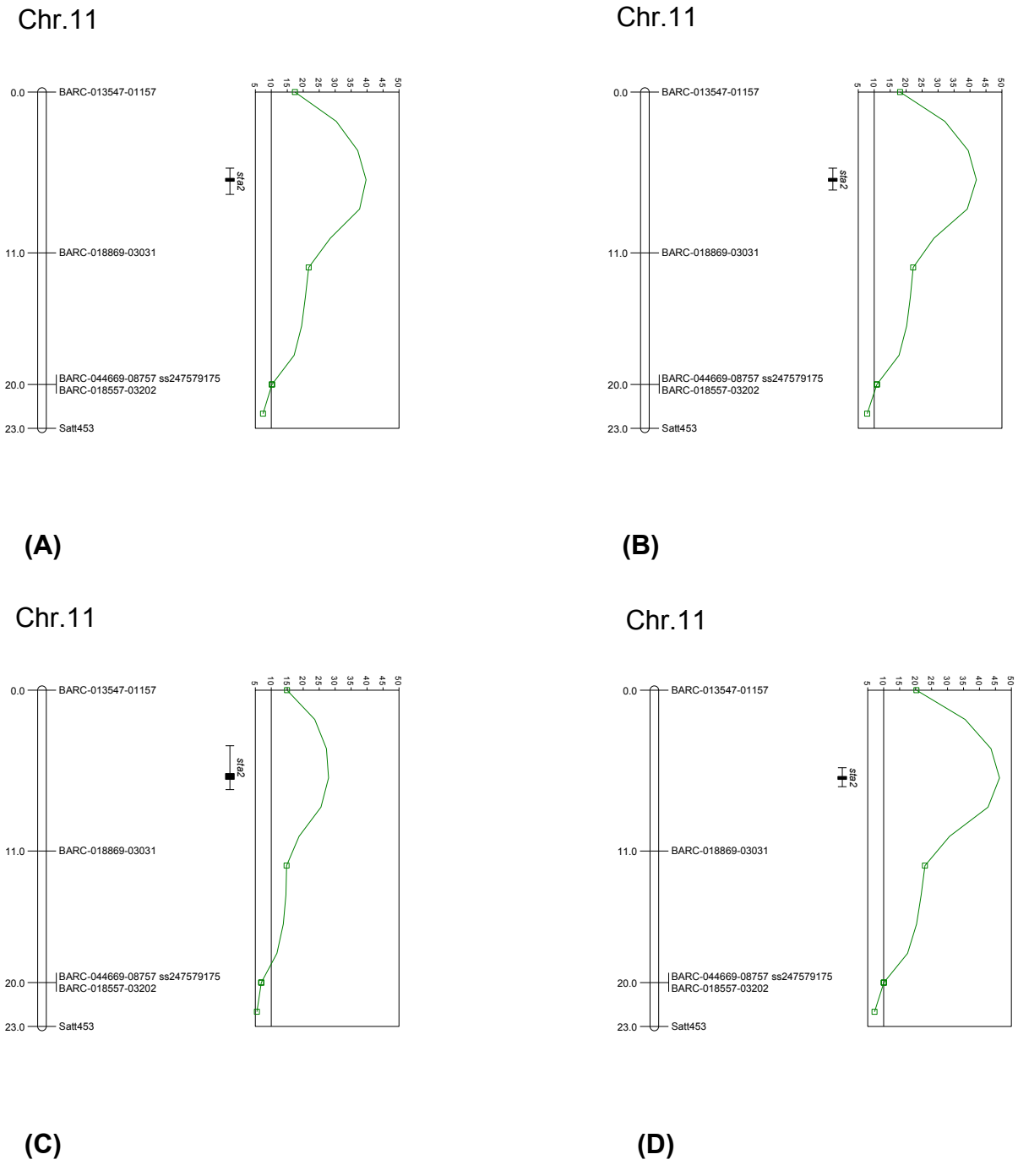


Figure 3. Composite interval mapping for seed stachyose QTL on chromosome 11 in 129 F_3-derived lines from Osage x V99-5089 evaluated across three environments: (A) in Fayetteville, AR in 2010; (B) in Argentina in 2011; (C) in Costa Rica in 2011; (D) combine data across 3 locations in 2 years.

OVERALL CONCLUSION

In this study, QTLs were identified for seed sucrose content in the population derived from MFS-553 x PI 243545 and QTLs/genes were identified for stachyose content in the population derived from Osage x V99-5089, using two sets of markers, multiple generations, and multiple environments. $F_{2:3}$ lines and $F_{3:4}$ lines were developed from MFS-553 x PI 243545 to map QTL for sucrose content. $F_{3:4}$ lines were developed from Osage x V99-5089 to map QTLs for stachyose content. $F_{2:3}$ and $F_{3:5}$ lines from MFS-553 x PI 243545 were evaluated in Fayetteville in 2009 and 2010, respectively. $F_{3:6}$ lines from MFS-553 x PI 243545 were evaluated in a randomized complete block design with two replications by two locations (Argentina and Costa Rica) in 2011. $F_{3:5}$ lines from Osage x V99-5089 were evaluated in Fayetteville in 2010, $F_{3:6}$ lines from Osage x V99-5089 were evaluated in a randomized complete block design with two replications by two locations (Argentina and Costa Rica) in 2011. The sucrose content in MFS-553 x PI 243545 fitted to a normal distribution whereas stachyose content in Osage x V99-5089 appeared to fit to a bimodal distribution.

Sucrose QTL mapping using SSR markers in $F_{2:3}$ lines from MFS-553 x PI 243545 revealed two QTLs on chromosome 5 and 16, respectively. Sucrose QTL mapping using SNP markers in $F_{3:4}$ lines from MFS-553 x PI 24354 revealed three QTLs on chromosome 5, 9, and 16, respectively. And sucrose QTLs identified from $F_{2:3}$ lines from MFS-553 x PI 24354 were confirmed and narrowed down by the F_3-derived lines from the same mapping population. Stachyose QTL mapping using SSR and SNP markers in $F_{3:4}$ lines from Osage x V99-5089 revealed that stachyose content was controlled by two QTLs/genes, the major QTL/gene was on chromosome 11

contributing the decrease of stachyose by 2.42%, the minor QTL/gene was on chromosome 10 contributing up to 1% of stachyose reduction. The Chi-square test further indicated that stachyose content was controlled by two independent genes.

These QTL/genes and associated markers for sucrose or stachyose content were stable in different genetic backgrounds and across different environments, they can be used for marker assisted selection in breeding soybean lines with desired sugar profile. To further reveal the genetic control of soybean seed sucrose and stachyose, more markers should be added in the putative QTL region to pinpoint the candidate genes.